



QY 241 AAGGAATACAGGCTCAGAGAACAGATGGAGGAGAGAGAGAAACCAATAGAGAAATC 300  
|||||  
Db 253 AAGGAATACAGGCTCAGAGAACAGATGGAGGAGAGAGAGAAACCAATAGAGAAATC 312  
|||||  
QY 301 CTTGGGAAAAACAACAGGAACCTACTAAGAAGAGTGAGAAGCAGTTACAGTGTGCTA 360  
|||||  
Db 313 CTTGGGAAAAACAACAGGNAACCTACTAGAAGAGTGAGAAGCAGTTACAGTGTGCTA 372  
|||||  
QY 361 ACACACTGCAATTAAGGTGCAATGCTTCCCTGGACCCAGCCCTGCGACCCCAACATCACC 420  
|||||  
Db 373 ACACACTGCAATTAAGGTGCAATGCTTCCCTGGACCCAGCCCTGCGACCCCAACATCACC 432  
|||||  
QY 421 CTGAAGGACCTGCATCTCTTTATCTCTTTATCTCTCTTTATCTCTCTTTATCTCTCT 480  
|||||  
Db 433 CTGAAGGACCTGCATCTCTTTATCTCTCTTTATCTCTCTTTATCTCTCTTTATCTCT 492  
|||||  
QY 481 GCAAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTTTTCACAGTACTAAGAT 540  
|||||  
Db 493 GCAAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTTTTCACAGTACTAAGAT 552  
|||||  
QY 541 GTTGAAGATGCCACTCTGACTGCTTTAAAGGAAATAGCAGACAGACATATGGAGGGC 600  
|||||  
Db 553 GTTGAAGATGCCACTCTGACTGCTTTAAAGGAAATAGCAGACAGACATATGGAGGGC 612  
|||||  
QY 601 TACTTGACACAGATAAAGATCCCTGATAGTCTAATGCAAGATTTGGCTTTCTG 660  
|||||  
Db 613 TACTTGACACAGATAAAGATCCCTGATAGTCTAATGCAAGATTTGGCTTTCTG 672  
|||||  
QY 661 AGTGGATTTCCACCAATCTCAAGGAAACAAATCAACTAAAACAAATTTGCCAGAAAT 720  
|||||  
Db 673 AGTGGATTTCCACCAATCTCAAGGAAACAAATCAACTAAAACAAATTTGCCAGAAAT 732  
|||||  
QY 721 CTTTCTCTAAAGTCAAACTGTTCAGTTGTTATTCAGAGGCCAGTGTAGCGTTCTCTAAA 780  
|||||  
Db 733 CTTTCTCTAAAGTCAAACTGTTCAGTTGTTATTCAGAGGCCAGTGTAGCGTTCTCTAAA 792  
|||||  
QY 781 CTGAATTAACCCNAGGATTTCAAGAAATGAATAAGCAAACTAAGAAGAACATGACCAAT 840  
|||||  
Db 793 CTGAATTAACCCNAGGATTTCAAGAAATGAATAAGCAAACTAAGAAGAACATGACCAAT 852  
|||||  
QY 841 GATGAAAAAGAACTGACCAATAGTCTGCATATTTATTTGAGTCTGAGGCCATCAGC 900  
|||||  
Db 853 GATGAAAAAGAACTGACCAATAGTCTGCATATTTATTTGAGTCTGAGGCCATCAGC 912  
|||||  
QY 901 CAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTGAAGATAACGAAGAACTGAGG 960  
|||||  
Db 913 CAGTCTAAGCAGATGAAGACATCTCTGCCAGTCTGTTTGAAGATAACGAAGAACTGAGG 972  
|||||  
QY 961 TACTCATTCGGATCTATCGAGAGGATGACCATGGTTCGGAATATTTTCATGTCACC 1020  
|||||  
Db 973 TACTCATTCGGATCTATCGAGAGGATGACCATGGTTCGGAATATTTTCATGTCACC 1032  
|||||  
QY 1021 AACGGCAGATTTCCATCTCGCTGAACCTTGACAATCTCGAGTGACAAATAGTAAACAC 1080  
|||||  
Db 1033 AACGGCAGATTTCCATCTCGCTGAACCTTGACAATCTCGAGTGACAAATAGTAAACAC 1092  
|||||  
QY 1081 CAGGATGTTTTTCGAAATTTGAGCCACTTGCCCTACCTTTAGTTCACCTGCTATTGAAAGT 1140  
|||||  
Db 1093 CAGGATGTTTTTCGAAATTTGAGCCACTTGCCCTACCTTTAGTTCACCTGCTATTGAAAGT 1152  
|||||  
QY 1141 CAGGTTATCGATCGAAGGGCTGTCAGAAAGTTTATTTACCTAAATGATGTCATG 1200  
|||||  
Db 1153 CAGGTTATCGATCGAAGGGCTGTCAGAAAGTTTATTTACCTAAATGATGTCATG 1212  
|||||  
QY 1201 TTTGGGAAGGATGCTGGCCAGATGATTTTACACTCACTCCAAAGGCCAGAGGTTTAT 1260  
|||||  
Db 1213 TTTGGGAAGGATGCTGGCCAGATGATTTTACACTCACTCCAAAGGCCAGAGGTTTAT 1272  
|||||  
QY 1261 TTGACATGCCCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCTCTGATTAAGGATGGC 1320  
|||||  
Db 1273 TTGACATGCCCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCTCTGATTAAGGATGGC 1332  
|||||  
QY 1321 TATTGTGACAGGCTTGTAAATAATTCAGCCCTGCGATTTGGGATGGTGGGATTTGCTCTGGA 1380  
|||||

Db 1333 TATTGTGACAGGCTTGTAAATAATTCAGCCCTGCGATTTGGGATGGGATTTGCTCTGGA 1392  
|||||  
QY 1381 AACAGTGGAGGAGTCCGTATATTTGCAGGAGGTGAGGTACTGGAGTATTTGGATTTGGA 1440  
|||||  
Db 1393 AACAGTGGAGGAGTCCGTATATTTGCAGGAGGTGAGGTACTGGAGTATTTGGAGTTGGA 1452  
|||||  
QY 1441 CAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTACTTAATCAGGATGTGGC 1500  
|||||  
Db 1453 CAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTACTTAATCAGGATGTGGC 1512  
|||||  
QY 1501 AATTCTGGCTGCTGATATAAGTTCTGTGACCAAGCATGCAATGCTTGTCTCTGTGGGTTT 1560  
|||||  
Db 1513 AATTCTGGCTGCTGATATAAGTTCTGTGACCAAGCATGCAATGCTTGTCTCTGTGGGTTT 1572  
|||||  
QY 1561 GATCTGCGACTCTGGGCAAGATCAATTTTCATGAATTTGATAAAGTGTATTAAGTGTCTCTCCCA 1620  
|||||  
Db 1573 GATCTGCGACTCTGGGCAAGATCAATTTTCATGAATTTGATAAAGTGTATTAAGTGTCTCTCCCA 1632  
|||||  
QY 1621 AACCAGACTCACTATATTTATTTCCAAAAGGTGAATGCCCTGCTTATTTTCAGCTTTGCAGAA 1680  
|||||  
Db 1633 AACCAGACTCACTATATTTATTTCCAAAAGGTGAATGCCCTGCTTATTTTCAGCTTTGCAGAA 1692  
|||||  
QY 1681 GTAGCCAAAAGAGGAGTTGAAGTGCCTATAGTGACAATCCAATPAATTCGACATGCTTCT 1740  
|||||  
Db 1693 GTAGCCAAAAGAGGAGTTGAAGTGCCTATAGTGACAATCCAATPAATTCGACATGCTTCT 1752  
|||||  
QY 1741 ATTGCCAACAGTGGAAACCATCCCTCATATGCACAGTGAATGAATGCCACCACA 1800  
|||||  
Db 1753 ATTGCCAACAGTGGAAACCATCCCTCATATGCACAGTGAATGAATGCCACCACA 1812  
|||||  
QY 1801 ATACATTTTATCTCAGCTTTTCAAAATACAAAGATGAAGAGTTCAAAATGCAGATAACA 1860  
|||||  
Db 1813 ATACATTTTATCTCAGCTTTTCAAAATACAAAGATGAAGAGTTCAAAATGCAGATAACA 1872  
|||||  
QY 1861 GTGAGGTGGACACAAGGGAGGACCAAACTGAAATTTCTACGCCCAAGAGGTTACGAA 1920  
|||||  
Db 1873 GTGAGGTGGACACAAGGGAGGACCAAACTGAAATTTCTACGCCCAAGAGGTTACGAA 1932  
|||||  
QY 1921 AATTAGTTAGTCCCATTAACACTTCTTCCAGAGCGGAAATCCTTTTGAGGATATTTCCC 1980  
|||||  
Db 1933 AATTAGTTAGTCCCATTAACACTTCTTCCAGAGCGGAAATCCTTTTGAGGATATTTCCC 1992  
|||||  
QY 1981 AAGAAAAACGCTTCCCGAAGTTTAAAGACATGATGTAACTCAACAAGAGAGAGCCAG 2040  
|||||  
Db 1993 AAGAAAAACGCTTCCCGAAGTTTAAAGACATGATGTAACTCAACAAGAGAGAGCCAG 2052  
|||||  
QY 2041 GAAGAGTGAATTTCCCTGGTAAATATTTACTCTTCCAAAAGACGCCAGTTGAGT 2100  
|||||  
Db 2053 GAAGAGTGAATTTCCCTGGTAAATATTTACTCTTCCAAAAGACGCCAGTTGAGT 2112  
|||||  
QY 2101 CTCATACCTTGGATTTGCAACTGGACATGGAGACATCACTTTGAAGGATACAATTTG 2160  
|||||  
Db 2113 CTCATACCTTGGATTTGCAACTGGACATGGAGACATCACTTTGAAGGATACAATTTG 2172  
|||||  
QY 2161 TCCAAGTCAGCCTTGTGTGATCATTTCTGATGAACCTCACAGCATGCTAAAAATAAAAT 2220  
|||||  
Db 2173 TCCAAGTCAGCCTTGTGTGATCATTTCTGATGAACCTCACAGCATGCTAAAAATAAAAT 2232  
|||||  
QY 2221 CAGCTATATAACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGTT 2280  
|||||  
Db 2233 CAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAGTT 2292  
|||||  
QY 2281 CATAAAGCATCTGCCAACACGCTTAGGAGTGTCTGAAAGATTTGAGAGTTGACATTTT 2340  
|||||  
Db 2293 CATAAAGCATCTGCCAACACGCTTAGGAGTGTCTGAAAGATTTGAGAGTTGACATTTT 2352  
|||||  
QY 2341 CCTGCAGTGTAGTAAAAAGTGAATGGTATGACAGGGTCAAGATCCACCTGGACTTTG 2400  
|||||  
Db 2353 CCTGCAGTGTAGTAAAAAGTGAATGGTATGACAGGGTCAAGATCCACCTGGACTTTG 2412  
|||||  
QY 2401 GAGACCACCAAGATTTAGAGTGGAAATTCACACCCAAAAACCAATAGCGGAAATGTG 2460  
|||||

Db 2413 GAGACCACAGCAAGATTAGAGTGGAAACTCACACCCAAAAAACCATTAGCGGAAATGTG 2472  
Qy 2461 ACAAAAGAAAGCCCCCATCTGTGATTTGTTCTCACTGTGAAGCCAGATGACAAAGAAAG 2520  
Db 2473 ACAAAAGAAAGCCCCCATCTGTGATTTGTTCTCACTGTGAAGCCAGATGACAAAGAAAG 2532  
Qy 2521 AAAATCACAGGGAAGAAAAAGAACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2580  
Db 2533 AAAATCACAGGGAAGAAAAAGAACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2592  
Qy 2581 GGCCTTACTGAAGTGTACTTGAAGAAAGCTGCAGCATTTACAGATGATTTACTTTGGGC 2640  
Db 2593 GGCCTTACTGAAGTGTACTTGAAGAAAGCTGCAGCATTTACAGATGATTTACTTTGGGC 2652  
Qy 2641 TTTTGGCCATGGGAGAAAAAAGTATTTCTTAGATTTCTTCAGCAAGAGAGTCAATG 2700  
Db 2653 TTTTGGCCATGGGAGAAAAAAGTATTTCTTAGATTTCTTCAGCAAGAGAGTCAATG 2712  
Qy 2701 AAGACAAATTTGGCTACTTCACTGATGACAAAGTACAGCAAGTACAGCAAGTACAGTAC 2760  
Db 2713 AAGACAAATTTGGCTACTTCACTGATGACAAAGTACAGCAAGTACAGCAAGTACAGTAC 2772  
Qy 2761 TTTGCAGATCCCTCAGATATGTAATTAATTTCTAAATAGCAAGTTTGGATTTACATCG 2820  
Db 2773 TTTGCAGATCCCTCAGATATGTAATTAATTTCTAAATAGCAAGTTTGGATTTACATCG 2832  
Qy 2821 CGGAAAGTCCCTCAGATATGTAATTAATTTCTAAATAGCAAGTTTGGATTTACATCG 2880  
Db 2833 CGGAAAGTCCCTCAGATATGTAATTAATTTCTAAATAGCAAGTTTGGATTTACATCG 2892  
Qy 2881 GATATGTTCCCTGAAGAAATTTGACACAGCTCAATTTACAAAGTGGCCCAATTTCTGAGGAT 2940  
Db 2893 GATATGTTCCCTGAAGAAATTTGACACAGCTCAATTTACAAAGTGGCCCAATTTCTGAGGAT 2952  
Qy 2941 ATGCAGTTTGGCTTCTCTTATTTTATTTATTTATCTCATGAGTGCAGTGCAGGCACTGA 3000  
Db 2953 ATGCAGTTTGGCTTCTCTTATTTTATTTATTTATCTCATGAGTGCAGTGCAGGCACTGA 3012  
Qy 3001 TCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTCTCTGTTCTGCTGACAGAAATC 3060  
Db 3013 TCTCAAGTCTTTGATGAGTTGATACAGATCAATCTGGTCTCTGTTCTGCTGACAGAAATC 3072  
Qy 3061 CGAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGGAGGATTTGACAGGCTG 3120  
Db 3073 CGAACACTGGCTACAGAAATTCAGAACTGCCGTTAAGTTTGGAGGATTTGACAGGCTG 3132  
Qy 3121 GAACACATGCTAATAATTTGCTCAAAATGCTTCCCTGCTGATATCAGCGAGCTAAATAAT 3180  
Db 3133 GAACACATGCTAATAATTTGCTCAAAATGCTTCCCTGCTGATATCAGCGAGCTAAATAAT 3192  
Qy 3181 ATTCACCAACTCAGGAATCTACTATGATCCCAACCTGCCAGCGTCACTAAAGTCTA 3240  
Db 3193 ATTCACCAACTCAGGAATCTACTATGATCCCAACCTGCCAGCGTCACTAAAGTCTA 3252  
Qy 3241 GTAACAACTGTAACCACTGAATGACAAATCCACAAAGCATATAGGACAAAAACAAA 3300  
Db 3253 GTAACAACTGTAACCACTGAATGACAAATCCACAAAGCATATAGGACAAAAACAAA 3312  
Qy 3301 TATAGGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAGTT 3360  
Db 3313 TATAGGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAGTT 3372  
Qy 3361 TCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAACCTAGGAAGTTTGTGGCTG 3420  
Db 3373 TCTCATGTGTTGGCCAGTTGGATGACATAAGAAAAACCTAGGAAGTTTGTGGCTG 3432  
Qy 3421 AATGACAACTTGACCACTAATAAGATGCTCAGACAGTGAAGGCTGTTCTCAGGGAC 3480  
Db 3433 AATGACAACTTGACCACTAATAAGATGCTCAGACAGTGAAGGCTGTTCTCAGGGAC 3492  
Qy 3481 TTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGCGCAAGAGAGTATCGAATC 3540  
Db 3493 TTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAATGCGCAAGAGAGTATCGAATC 3552

Qy 3541 CGTTTCTTCATATCATGAGCTGCAGGAATGGAGGGCTTATCGAGACAAATTTGAAGTAG 3600  
Db 3553 CGTTTCTTCATATCATGAGCTGCAGGAATGGAGGGCTTATCGAGACAAATTTGAAGTAG 3612

## RESULT 2

US-09-986-552-21  
; Sequence 21, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 21  
; LENGTH: 3621  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-986-552-21

Query Match 99.5%; Score 3580.8; DB 10; Length 3621;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 3588; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ATGGAGACAGACACACTCCTGCTATGGTGGTCTGCTGCTGGTTCAGGTTCCAGTTCCACTGGT 60  
Db 13 ATGGAGACAGACACACTCCTGCTATGGTGGTCTGCTGCTGGTTCAGGTTCCAGTTCCACTGGT 72  
Qy 61 GACGAAGATCAGGTAGATCGCGGTTAATCGACGGTAACTAGCGGAGATCAATACCAT 120  
Db 73 GACGAAGATCAGGTAGATCGCGGTTAATCGACGGTAACTAGCGGAGATCAATACCAT 132  
Qy 121 GTTTGTTGTTGATTCTCTATAGACAAATATGCTGGAAGTCTCTTCAGAAATCGGCTTTGT 180  
Db 133 GTTTGTTGTTGATTCTCTATAGACAAATATGCTGGAAGTCTCTTCAGAAATCGGCTTTGT 192  
Qy 181 CTGCCCATCCCATTCAGCTTCTTACACCTGSGTGAATGGCACAGATCTTGAACACTG 240  
Db 193 CTGCCCATCCCATTCAGCTTCTTACACCTGSGTGAATGGCACAGATCTTGAACACTG 252  
Qy 241 AAGGAACACTACAGAGTCTAGAAACAGATGGAGGAGGAGAGAGAGAGAGAGAGAGAGATC 300  
Db 253 AAGGAACACTACAGAGTCTAGAAACAGATGGAGGAGGAGAGAGAGAGAGAGAGAGATC 312  
Qy 301 CTTGGGAAAAACACAGCGAACCCTACTAAGAGAGTGAAGAGCAGTGTAGAGTGTGGTA 360  
Db 313 CTTGGGAAAAACACAGCGAACCCTACTAAGAGAGTGAAGAGCAGTGTAGAGTGTGGTA 372  
Qy 361 ACACACTGCTAATAAGTGGCAATGCTGCTGAGCCAGCCCTGCCAGCCATCAACATCACC 420  
Db 373 ACACACTGCTAATAAGTGGCAATGCTGCTGAGCCAGCCCTGCCAGCCATCAACATCACC 432  
Qy 421 CTGAAGGACCTGCCATCTCTTTATCTCTTTTCAATCTGCGCAGTGACATTTTCAATGTT 480  
Db 433 CTGAAGGACCTGCCATCTCTTTATCTCTTTTCAATCTGCGCAGTGACATTTTCAATGTT 492  
Qy 481 GCAAAACAAAAAACCCTTCTACCAATGCTCAGTGTGTTGTTTTCAGACAGTACTAAGAT 540  
Db 493 GCAAAACAAAAAACCCTTCTACCAATGCTCAGTGTGTTGTTTTCAGACAGTACTAAGAT 552  
Qy 541 GTTGAAGATGCCACTCTCGACTGCTTAAGGAAATACGACACACAGTATGGAGGGC 600  
Db 553 GTTGAAGATGCCACTCTCGACTGCTTAAGGAAATACGACACACAGTATGGAGGGC 612  
Qy 601 TACTTGACAACAGATAAAGAAGTCCCTGGATTAGTCTAATGCAAGATTTGGCTTTTCCTG 660

Db 613 TACTGTGACACAGATAAAGAGTCCCTGGATTAGTGTAAATGCAAGATTGGCTTCTCTG 672  
Qy 661 AGTGGATTCCACCAACATTCACGGAACAAATCAACTAAACAAAATGGCCAGAAAAT 720  
Db 673 AGTGGATTCCACCAACATTCACGGAACAAATCAACTAAACAAAATGGCCAGAAAAT 732  
Qy 721 CTTTCCTCTAAGTCAAACTGTTCCAGTTGTATTACAGAGCCAGTGTAGCGCTTCTAAAA 780  
Db 733 CTTTCCTCTAAGTCAAACTGTTCCAGTTGTATTACAGAGCCAGTGTAGCGCTTCTAAAA 792  
Qy 781 CTGAATTAACCCCAAGGATTTTCAAGATTAAGTAAGCAAACTAAGAAACATGACCAAT 840  
Db 793 CTGAATTAACCCCAAGGATTTTCAAGATTAAGTAAGCAAACTAAGAAACATGACCAAT 852  
Qy 841 GATGGAAGAACTGACCAATAGTCTGCAATTAATTAATAGGATCTGAGCGCCATCAGC 900  
Db 853 GATGGAAGAACTGACCAATAGTCTGCAATTAATTAATAGGATCTGAGCGCCATCAGC 912  
Qy 901 CAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATTAAGCAAACTGAGG 960  
Db 913 CAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGAAGATTAAGCAAACTGAGG 972  
Qy 961 TACTCATTTGGCATCTATCGAGAGGATGCAACCATGGGTTTCGGAATATTTTCATTTGCACC 1020  
Db 973 TACTCATTTGGCATCTATCGAGAGGATGCAACCATGGGTTTCGGAATATTTTCATTTGCACC 1032  
Qy 1021 AACGGCAGATTTCCATCTCGCTGAACCTTGACAATCTCGAGTGCACATAGTAACACAC 1080  
Db 1033 AACGGCAGATTTCCATCTCGCTGAACCTTGACAATCTCGAGTGCACATAGTAACACAC 1092  
Qy 1081 CAGGATGTTTTTGGAAATTTGACCACTTGCCCTACCTTTAGTTTCACTCTATTTGAAAGT 1140  
Db 1093 CAGGATGTTTTTGGAAATTTGACCACTTGCCCTACCTTTAGTTTCACTCTATTTGAAAGT 1152  
Qy 1141 CAGCTTCATCGCATCGAAGGGCTGTCCCAAGATTTTATTTACCTAAATGATGATGTCATG 1200  
Db 1153 CAGCTTCATCGCATCGAAGGGCTGTCCCAAGATTTTATTTACCTAAATGATGATGTCATG 1212  
Qy 1201 TTTGGGAAGGATCTGTGGCCAGATGATTTTACAGTCACTTCCAAAGGCCAGAAAGTTTAT 1260  
Db 1213 TTTGGGAAGGATCTGTGGCCAGATGATTTTACAGTCACTTCCAAAGGCCAGAAAGTTTAT 1272  
Qy 1261 TTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGTTCTCTGGAATTAAGGATGSC 1320  
Db 1273 TTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGTTCTCTGGAATTAAGGATGSC 1332  
Qy 1321 TATTGTGACAAGCTTTGTATAATTAATCAGCTTCGATTTGGGATGGGATGCTCTGGA 1380  
Db 1333 TATTGTGACAAGCTTTGTATAATTAATCAGCTTCGATTTGGGATGGGATGCTCTGGA 1392  
Qy 1381 AACAGTGGAGGAGTCGCTATATTGACAGAGTGGAGGTACTGGAGTATTTGGATTTGGA 1440  
Db 1393 AACAGTGGAGGAGTCGCTATATTGACAGAGTGGAGGTACTGGAGTATTTGGATTTGGA 1452  
Qy 1441 CAGCCCTGCGACTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTAATCAGGATGTCGG 1500  
Db 1453 CAGCCCTGCGACTTTGGTGGAGGAATAAACAGTGTCTCTTACTGTAATCAGGATGTCGG 1512  
Qy 1501 AATTCTGCTGCTGTGATAAGTTCTGTGCCAAGCATGCAATGCTTCTGCTGTGGGTTT 1560  
Db 1513 AATTCTGCTGCTGTGATAAGTTCTGTGCCAAGCATGCAATGCTTCTGCTGTGGGTTT 1572  
Qy 1561 GATGCTGGGACTGTGGCAAGATCATTTTCATGAATTTGTAAGTCACTCTTCTCCCA 1620  
Db 1573 GATGCTGGGACTGTGGCAAGATCATTTTCATGAATTTGTAAGTCACTCTTCTCCCA 1632  
Qy 1621 AACAGACTCACTATATTATTTCCAAAGGTGAATGCCCTTATTTTTCAGCTTTTGCAGAA 1680  
Db 1633 AACAGACTCACTATATTATTTCCAAAGGTGAATGCCCTTATTTTTCAGCTTTTGCAGAA 1692  
Qy 1681 GTAGCCAAAGAGGAGTTGAAGTGGCTTATAGTGACAATCCAATTAATTCGACATGCTCT 1740

Db 1693 GTAGCCAAAGAGGAGTTGAAGGTGCCTATAGTGACAATCCAATTAATTCGACATGCTTCT 1752  
Qy 1741 ATTGCCAACAGTGGAAAAACCATCCACCTCAATTAATGCAGTGGAAATGAATGCCACCACA 1800  
Db 1753 ATTGCCAACAGTGGAAAAACCATCCACCTCAATTAATGCAGTGGAAATGAATGCCACCACA 1812  
Qy 1801 ATACATTTTAAATCTCACGTTTCAAAATACAAACGATGAAGAGTTTCAAAATGCAGATAACA 1860  
Db 1813 ATACATTTTAAATCTCACGTTTCAAAATACAAACGATGAAGAGTTTCAAAATGCAGATAACA 1872  
Qy 1861 GTGGAGTGGACACAAAGGAGGACCAAAACCTGAAATCTACGGCCCAAGGGTTACGAA 1920  
Db 1873 GTGGAGTGGACACAAAGGAGGACCAAAACCTGAAATCTACGGCCCAAGGGTTACGAA 1932  
Qy 1921 AATTTAGTTAGTCCCATTAACACTTCTCCAGAGGGGAAAACTCTTTTGGAGATATTCCC 1980  
Db 1933 AATTTAGTTAGTCCCATTAACACTTCTCCAGAGGGGAAAACTCTTTTGGAGATATTCCC 1992  
Qy 1981 AAAGAAAAACCGTTCCCGAAGTTTAAAGAGACATGATGTTAACTCAACAAGGAGAGCCAG 2040  
Db 1993 AAAGAAAAACCGTTCCCGAAGTTTAAAGAGACATGATGTTAACTCAACAAGGAGAGCCAG 2052  
Qy 2041 GAAAGAGTGAATAATTTCCCTCGTAAATATTTCACTCTCCAAAGAGCGCCAGTTGAGT 2100  
Db 2053 GAAAGAGTGAATAATTTCCCTCGTAAATATTTCACTCTCCAAAGAGCGCCAGTTGAGT 2112  
Qy 2101 CTCATACCTTTGGATTTGCAACTGGACATGGACATCACTTTTGAAGAGATACAATTTG 2160  
Db 2113 CTCATACCTTTGGATTTGCAACTGGACATGGACATCACTTTTGAAGAGATACAATTTG 2172  
Qy 2161 TCCAGTCAAGCTTGTCTGAGATCAATTTCTGATGAACCTCACAGCATGCTAAAAATAAAT 2220  
Db 2173 TCCAGTCAAGCTTGTCTGAGATCAATTTCTGATGAACCTCACAGCATGCTAAAAATAAAT 2232  
Qy 2221 CAAGCTATAATAACAGATGAACAAATGACAGTTTGTGGCTCCACAGGAAAAACAGGTT 2280  
Db 2233 CAAGCTATAATAACAGATGAACAAATGACAGTTTGTGGCTCCACAGGAAAAACAGGTT 2292  
Qy 2281 CATAAAAGCATCTTGGCAACAGCTTAGGAGTGTCTGAAAGATTCAGAGAGTTGACTTTT 2340  
Db 2293 CATAAAAGCATCTTGGCAACAGCTTAGGAGTGTCTGAAAGATTCAGAGAGTTGACTTTT 2352  
Qy 2341 CCTGCAGTGTGTTAAAGTGAATGTCATGACAGGGTCAGAAATCCACCCCTGGAGTTG 2400  
Db 2353 CCTGCAGTGTGTTAAAGTGAATGTCATGACAGGGTCAGAAATCCACCCCTGGAGTTG 2412  
Qy 2401 GAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAACCATAGCGGAAATGTG 2460  
Db 2413 GAGACCACAGCAAGATTTAGAGTGGAAACTCACACCCAAAAACCATAGCGGAAATGTG 2472  
Qy 2461 ACAAAGAAAGCCCTCATCTCTGATTGTTCCACTGGAAAGCCAGATGACAAAGAAAAAG 2520  
Db 2473 ACAAAGAAAGCCCTCATCTCTGATTGTTCCACTGGAAAGCCAGATGACAAAGAAAAAG 2532  
Qy 2521 AAAATCACAGGAAAGAAAGAGACAGTAGAATGGAGGAAAAATGCTGAAATCAACATA 2580  
Db 2533 AAAATCACAGGAAAGAAAGAGACAGTAGAATGGAGGAAAAATGCTGAAATCAACATA 2592  
Qy 2581 GGGCTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACACAGATAGTTACTTTGGC 2640  
Db 2593 GGGCTTACTGAAGTGTACTTTGGAAGAAAGCTGCAGCATTTACACAGATAGTTACTTTGGC 2652  
Qy 2641 TTTTGGCCATGGGAAAAAAAAGTATTTCTCTAGATCTTCTCGAACCAAGAGATCATTTG 2700  
Db 2653 TTTTGGCCATGGGAAAAAAAAGTATTTCTAGATCTTCTCGAACCAAGAGATCATTTG 2712  
Qy 2701 AAGACAAATGGCTTACTTCACTGATACAGAAATAGAGCCAGATACAGAGAGATACA 2760  
Db 2713 AAGACAAATGGCTTACTTCACTGATACAGAAATAGAGCCAGATACAGAGAGATACA 2772  
Qy 2761 TTTGCGAGTTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCG 2820  
Db 2773 TTTGCGAGTTTCCCTCAGATATGTAATAAATTTCTAAATAGCAAGTTTGGATTCACATCG 2832



QY 2821 CGGAAAGTCCCTGCTCACATGCTCCTCACATGATTCGCGGATGTTTATGCAAGAACTGCAA 2880  
Db 2833 CGGAAAGTCCCTGCTCACATGCTCCTCACATGATTCGCGGATGTTTATGCAAGAACTGCAA 2892  
QY 2881 GATATGTTCCCTGAAGAAATTCACAGACGTCATTTACAAAGTGGCCATTCAGGAT 2940  
Db 2893 GATATGTTCCCTGAAGAAATTCACAGACGTCATTTACAAAGTGGCCATTCAGGAT 2952  
QY 2941 ATGCAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCAGTGCGGACCTGGAATATA 3000  
Db 2953 ATGCAGTTTGCCTTCTCTTATTTTATATCTCATGAGTGCAGTGCGGACCTGGAATATA 3012  
QY 3001 TCTCAAGTCTTGTATGAAGTTGTATACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATC 3060  
Db 3013 TCTCAAGTCTTGTATGAAGTTGTATACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATC 3072  
QY 3061 CGAACACTGGCTACCGAATTCACGAACCTGCCGTTAAGTTTGCAGGATTTGACAGTCTG 3120  
Db 3073 CGAACACTGGCTACCGAATTCACGAACCTGCCGTTAAGTTTGCAGGATTTGACAGTCTG 3132  
QY 3121 GAACACATGCTAATAATTCCTCAAAATGCTTCCTGCTGATATCAGCAGCTTAAATAT 3180  
Db 3133 GAACACATGCTAATAATTCCTCAAAATGCTTCCTGCTGATATCAGCAGCTTAAATAT 3192  
QY 3181 ATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCACCGTCACTAAAGTCTA 3240  
Db 3193 ATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCACCGTCACTAAAGTCTA 3252  
QY 3241 GTAAACAACTGTAACACGTAAGTACGACAAATCCCAAGCATATAGGACAAACAAA 3300  
Db 3253 GTAAACAACTGTAACACGTAAGTACGACAAATCCCAAGCATATAGGACAAACAAA 3312  
QY 3301 TATAGTTTGAATCATGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAGCTT 3360  
Db 3313 TATAGTTTGAATCATGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAGCTT 3372  
QY 3361 TCTCATGTGTTGGCCAGTTGGATGACATAAGAAACAAACCTAGGAAGTTTGTGCTG 3420  
Db 3373 TCTCATGTGTTGGCCAGTTGGATGACATAAGAAACAAACCTAGGAAGTTTGTGCTG 3432  
QY 3421 AATGACAACTGACCAACATCAATAAGATGCTCAGACAGTGAAGCTGTCTCAGGAC 3480  
Db 3433 AATGACAACTGACCAACATCAATAAGATGCTCAGACAGTGAAGCTGTCTCAGGAC 3492  
QY 3481 TTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAAGAGATATCGAAAC 3540  
Db 3493 TTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACCTGCCAAGAGATATCGAAAC 3552  
QY 3541 CGTTTCTCTCATATGATGAGCTGCAGGAATGGAGGCTTATCGAGACAAATTTGAAGTAG 3600  
Db 3553 CGTTTCTCTCATATGATGAGCTGCAGGAATGGAGGCTTATCGAGACAAATTTGAAGTAG 3612

## RESULT 3

US-09-895-072-20  
; Sequence 20, Application US/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCNT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 20  
; LENGTH: 3783  
; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-895-072-20

## Query Match

Best Local Similarity 96.6%; Score 3476.8; DB 10; Length 3783;

Matches 3484; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 103 AGCCAGATCAATACCATGTTTGTTCATTCCTATAGACAATATTCCTGGAAAGTCC 162  
Db 139 AGCCAGATCAATACCATGTTTGTTCATTCCTATAGACAATATTCCTGGAAAGTCC 198  
QY 163 TTTCAGAAATCGGCTTTGTCTGCCCATGCGGATTTGACAGTGTGTTTACACCTGGTGAATGGC 222  
Db 199 TTTCAGAAATCGGCTTTGTCTGCCCATGCGGATTTGACAGTGTGTTTACACCTGGTGAATGGC 258  
QY 223 ACAGATCTTGAATCTACTGAAGAACTACAGAGAGTTCAGAGACAGATGGAGAGGAGCAG 282  
Db 259 ACAGATCTTGAATCTACTGAAGAACTACAGAGAGTTCAGAGACAGATGGAGAGGAGCAG 318  
QY 283 AAAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAAGAGAGTGAGAG 342  
Db 319 AAAGCAATGAGAGAAATCCTTGGGAAACACACGGAACCTACTAAAGAGAGTGAGAG 378  
QY 343 CAGTTAGAGTGTTCCTAACACACTGCATTAAGGTGCCAATGCTTGTCTTGACCCAGCC 402  
Db 379 CAGTTAGAGTGTTCCTAACACACTGCATTAAGGTGCCAATGCTTGTCTTGACCCAGCC 438  
QY 403 CTGCCAGCAACATCACCCTGAAGACCTGCCATCTCTTTATCTCTTTTCATTCCTGCC 462  
Db 439 CTGCCAGCAACATCACCCTGAAGACCTGCCATCTCTTTATCTCTTTTCATTCCTGCC 498  
QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCAGTTGTTGTT 522  
Db 499 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTACCAATGCTCAGTTGTTGTT 558  
QY 523 TTTGACAGTACTAAGGATGTTTGAAGATGCCACTCTGGACTGCTTTAAAGGAAATAGCAGA 582  
Db 559 TTTGACAGTACTAAGGATGTTTGAAGATGCCACTCTGGACTGCTTTAAAGGAAATAGCAGA 618  
QY 583 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATAGTGTGTAATG 642  
Db 619 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATAGTGTGTAATG 678  
QY 643 CAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAA 702  
Db 679 CAAGATTTGGCTTTCTCTGAGTGGATTTCCACCAACATTCAGAGAAACAAATCAACTAAA 738  
QY 703 ACAAAATGCCAGAAATCTTCT 762  
Db 739 ACAAAATGCCAGAAATCTTCT 798  
QY 763 AGTGTAGCGCTTCTAAACCTGAATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 822  
Db 799 AGTGTAGCGCTTCTAAACCTGAATTAACCCCAAGGATTTTCAAGATTTGAATTAAGCAACT 858  
QY 823 AAGAGAAACATGACCATGATTCATGAAAGAACTGACATAAGTCCCTGCATATTTATATGG 882  
Db 859 AAGAGAAACATGACCATGATTCATGAAAGAACTGACATAAGTCCCTGCATATTTATATGG 918  
QY 883 GATCTGAGCGCCATCAGCAGTCTTAAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTAA 942  
Db 919 GATCTGAGCGCCATCAGCAGTCTTAAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTAA 978  
QY 943 GATAACGAAACTGAGGTACTCTATCGAGAGGATTCATCGAGAGGATTCGACATCTCTCGG 1002  
Db 979 GATAACGAAACTGAGGTACTCTATCGAGAGGATTCATCGAGAGGATTCGACATCTCTCGG 1038  
QY 1003 AATATTTTCTATGTCACCAACGGGAGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1062  
Db 1039 AATATTTTCTATGTCACCAACGGGAGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1098  
QY 1063 GTGACAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTTACCTTACCT 1122

Db 1099 GTGCAATAGTAACACACCAGGATGTTTTTGGAAATTTGAGCCACTTGCCTACCTTTAGT 1158  
QY 1123 TCACCTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCTGTCCCAAGAGTTTATTAC 1182  
Db 1159 TCACCTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCTGTCCCAAGAGTTTATTAC 1218  
QY 1183 CTAATGATGATGTCATGTTTGGGAAGGATCTCTGGCCAGATGATTTTACAGTCACTCC 1242  
Db 1219 CTAATGATGATGTCATGTTTGGGAAGGATGTCGGCCAGATGATTTTACAGTCACTCC 1278  
QY 1243 AAAGCCAGAGGTTTATTGACATGGCCTGTGCCAAACTGTGCCAGGCTGCCAGGT 1302  
Db 1279 AAAGCCAGAGGTTTATTGACATGGCCTGTGCCAAACTGTGCCAGGCTGCCAGGT 1338  
QY 1303 TCCTGGATTAAAGATGCTATTGTGACAAAGGCTTGTAAATAATTCACGCTGCCATGGGAT 1362  
Db 1339 TCCTGGATTAAAGATGCTATTGTGACAAAGGCTTGTAAATAATTCACGCTGCCATGGGAT 1398  
QY 1363 GTTGGGATTTGCTCTGGAACAGTGGAGGAGTGCCTATATTGACAGAGTGGAGTACT 1422  
Db 1399 GTTGGGATTTGCTCTGGAACAGTGGAGGAGTGCCTATATTGACAGAGTGGAGTACT 1458  
QY 1423 GGAGATTTGAGCTTGGACAGCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1482  
Db 1459 GGAGATTTGAGCTTGGACAGCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1518  
QY 1483 TGTAAATCAGGATGTGCGAATTCCTGGCTCGCTGATAAGTCTGTGACCAAGCATGCCAAT 1542  
Db 1519 TGTAAATCAGGATGTGCGAATTCCTGGCTCGCTGATAAGTCTGTGACCAAGCATGCCAAT 1578  
QY 1543 GTCTTGTCTGTGGTTTGATGCTGGCGACTGTGGCAAGATCAATTTTCATGAATGTAT 1602  
Db 1579 GTCTTGTCTGTGGTTTGATGCTGGCGACTGTGGCAAGATCAATTTTCATGAATGTAT 1638  
QY 1603 AAAGTATCCTTCTCCCAACACAGACTCACTATATATTTCAAAAGTGAATGCTGCT 1662  
Db 1639 AAAGTATCCTTCTCCCAACACAGACTCACTATATATTTCAAAAGTGAATGCTGCT 1698  
QY 1663 TATTTACGCTTTGCAGAGTAGCCAAAAGAGAGTTGAAGTGGCTATAGTGACAATCCA 1722  
Db 1699 TATTTACGCTTTGCAGAGTAGCCAAAAGAGAGTTGAAGTGGCTATAGTGACAATCCA 1758  
QY 1723 ATAATTCGACATGCTTCTATTGCGCAACAGTGGAACCATCCACCTCATATGACAGT 1782  
Db 1759 ATAATTCGACATGCTTCTATTGCGCAACAGTGGAACCATCCACCTCATATGACAGT 1818  
QY 1783 GGAATGAATGCCACCAATACATTTTAATCTCAGCTTTCAAAATACAAACGATGAAG 1842  
Db 1819 GGAATGAATGCCACCAATACATTTTAATCTCAGCTTTCAAAATACAAACGATGAAG 1878  
QY 1843 TTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGGAGGACCAAACTGAATTTCTACG 1902  
Db 1879 TTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGGAGGACCAAACTGAATTTCTACG 1938  
QY 1903 GCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATACACATTCCTCCAGAGGCGGAAATC 1962  
Db 1939 GCCAGAGAGGTTACGAAAATTTAGTTAGTCCCATACACATTCCTCCAGAGGCGGAAATC 1998  
QY 1963 CTTTTTGGAGTATTTCCCAAGAAAAAGCTTCCCGAAGTTTAAGACACATGATTTAAC 2022  
Db 1999 CTTTTTGGAGTATTTCCCAAGAAAAAGCTTCCCGAAGTTTAAGACACATGATTTAAC 2058  
QY 2023 TCACAGGAGGAGGCCAGGAGGTTGAAAATTTCCCTGGTGAATATTTTCACCTCCCTTCCA 2082  
Db 2059 TCACAGGAGGAGGCCAGGAGGTTGAAAATTTCCCTGGTGAATATTTTCACCTCCCTTCCA 2118  
QY 2083 AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACT 2142  
Db 2119 AAAGACCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACT 2178  
QY 2143 TTGAAAGATACAAATTTGTCCAAAGTACAGCTTGTGAGATCAATTTCTGATGAATCACAG 2202  
Db 2179 TTGAAAGATACAAATTTGTCCAAAGTACAGCTTGTGAGATCAATTTCTGATGAATCACAG 2238

QY 2203 CATGCTAAATAAAAATCAAGCTATAATACAGATGAAACAAATACAGTTTGGTGGCT 2262  
Db 2239 CATGCTAAATAAAAATCAAGCTATAATACAGATGAAACAAATACAGTTTGGTGGCT 2298  
QY 2263 CCACAGGAAAAACAGGTTTCATAAAGCATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGA 2322  
Db 2299 CCACAGGAAAAACAGGTTTCATAAAGCATCTTGGCCAAACAGCTTAGGAGTGTCTGAAAGA 2358  
QY 2323 TTGCAGAGGTTGACATTTTCCCTGAGTGTAAAGTGAATGTCATGACACAGGTCAG 2382  
Db 2359 TTGCAGAGGTTGACATTTTCCCTGAGTGTAAAGTGAATGTCATGACACAGGTCAG 2418  
QY 2383 AATCCACCCCTGGACTTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCACAAAA 2442  
Db 2419 AATCCACCCCTGGACTTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCACACCACAAAA 2478  
QY 2443 ACCATAGCGGAAATGTGACAAAAAGAAAGCCCCCATCTCTGATTGTTCACCTGGAAGC 2502  
Db 2479 ACCATAGCGGAAATGTGACAAAAAGAAAGCCCCCATCTCTGATTGTTCACCTGGAAGC 2538  
QY 2503 CAGATGACAAAAAGAAAGAAATCACAGGGAAGAAAAAGAGACAGTACAGTGGAGGAA 2562  
Db 2539 CAGATGACAAAAAGAAAGAAATCACAGGGAAGAAAAAGAGACAGTACAGTGGAGGAA 2598  
QY 2563 AATGCTGAAATACATAGGCGTTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTAC 2622  
Db 2599 AATGCTGAAATACATAGGCGTTTACTGAAAGTGTACTTGGAAAGAAAGCTGCAGCATTAC 2658  
QY 2623 ACAGATAGTTACTTGGCGTTTGGCCATGGGAGAAAAAGATTTTCCTAGATCTCTC 2682  
Db 2659 ACAGATAGTTACTTGGCGTTTGGCCATGGGAGAAAAAGATTTTCCTAGATCTCTC 2718  
QY 2683 GACGAAGAGAGTCAATGGAAGACACATTTGGCTTACTTCACTCATAGCAGAAATAGGCC 2742  
Db 2719 GACGAAGAGTCAATGGAAGACACATTTGGCTTACTTCACTCATAGCAGAAATAGGCC 2778  
QY 2743 AGATACAGAGAGATACATTTGCGAGATTCCTCAGATATGTAATAATAAATTTCTAAATAGC 2802  
Db 2779 AGGCACTAAAGATACATTTGCGAGATTCCTCAGATATGTAATAATAAATTTCTAAATAGC 2838  
QY 2803 AAGTTTGGATTACATCGCGGAAGTCCCTGCTCAGATGCTCAGATGATGATGACCGGAT 2862  
Db 2839 AAGTTTGGATTACATCGCGGAAGTCCCTGCTCAGATGCTCAGATGATGATGACCGGAT 2898  
QY 2863 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAGACGTCATTTACAAA 2922  
Db 2899 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAAGATTTGACAAGACGTCATTTACAAA 2958  
QY 2923 GTGCGCATTCGAGGATGATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGA 2982  
Db 2959 GTGCGCATTCGAGGATGATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGA 3018  
QY 2983 GTGCGCATTCGAGGATGATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGA 3042  
Db 3019 GTGCGCATTCGAGGATGATGAGTTTGGCTTCTCTTATTTTATCTCATGAGTGA 3078  
QY 3043 TTGCTGACAGAAATCCGAACACTGGCTACGAGAAATTCAGAACTGCGGTTAAGTTTG 3102  
Db 3079 TTGCTGACAGAAATCCGAACACTGGCTACGAGAAATTCAGAACTGCGGTTAAGTTTG 3138  
QY 3103 CAGGATTTGACAGGTTGGAACACATGCTATAATAATGCTCAAAAATGCTTCCTGCTGAT 3162  
Db 3139 CAGGATTTGACAGGTTGGAACACATGCTATAATAATGCTCAAAAATGCTTCCTGCTGAT 3198  
QY 3163 ATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGATCCCAACTGCCA 3222  
Db 3199 ATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGATCCCAACTGCCA 3258  
QY 3223 CCGGTCACTAAAGTCTAGTAACAACTGTAAACAGTAACTGACAAAATCCACAAAGCA 3282  
Db 3259 CCGGTCACTAAAGTCTAGTAACAACTGTAAACAGTAACTGACAAAATCCACAAAGCA 3318



[illegible]

RESULT 5  
US-09-895-072-4  
; Sequence 4, Application US/09895072  
; Patent No. US2002025550A1

GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 5597  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-895-072-4

Query Match 96.3%; Score 3465.6; DB 10; Length 5597;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 3477; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY	103	AGCCGAGATCAATACCATGTTTGGTGGATTCCTATAGAGACAATATTCCTGGAAGTCC	162
DB	297	AGCCGAGATCAATACCATGTTTGGTGGATTCCTATAGAGACAATATTCCTGGAAGTCC	356
QY	163	TTTCAGAAATCGGCTTTGCTGCCATGCCGATTGAGTTGTTTACACCTGGTGAATGCC	222
DB	357	TTTCAGAAATCGGCTTTGCTGCCATGCCGATTGAGTTGTTTACACCTGGTGAATGCC	416
QY	223	ACAGATCTTGAATCTACTAAGGAACCTACAGAGCTCAGAGACAGATGGAGGAGGAGCAG	282
DB	417	ACAGATCTTGAATCTACTAAGGAACCTACAGAGCTCAGAGACAGATGGAGGAGGAGCAG	476
QY	283	AAAGCAATCAGAAATCCTTGGGAAACACACAGCAACCTACTAAGAGAGTGAAGAG	342
DB	477	AAAGCAATCAGAAATCCTTGGGAAACACACAGCAACCTACTAAGAGAGTGAAGAG	536
QY	343	CAGTTAGATGTTGCTAACACACTGCATTAAGGTGCCAATGCTTGTCTGACCCAGCC	402
DB	537	CAGTTAGATGTTGCTAACACACTGCATTAAGGTGCCAATGCTTGTCTGACCCAGCC	596
QY	403	CTGCCAGCAACATCACCCTGAAGACCTGCCATCTCTTTAFCCTTTTCAATCTGCC	462
DB	597	CTGCCAGCAACATCACCCTGAAGACCTGCCATCTCTTTAFCCTTTTCAATCTGCC	656
QY	463	AGTGACATTTTCAATGTTGCCAAACCAACCAACCTCTTACCAATGCTCAGTTGTTGT	522
DB	657	AGTGACATTTTCAATGTTGCCAAACCAACCAACCTCTTACCAATGCTCAGTTGTTGT	716
QY	523	TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTTAAAGGAAATAGCAGA	582
DB	717	TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTTAAAGGAAATAGCAGA	776
QY	583	CAGACAGATGAGGGGCTACTTGACACACAGATAAAGAGTCCCTGGATTAGTCTAATG	642
DB	777	CAGACAGATGAGGGGCTACTTGACACACAGATAAAGAGTCCCTGGATTAGTCTAATG	836
QY	643	CAAGATTTGGCTTTCTCCTGAGTGGATTTCCCAACATTCAGGAAACAAATCAACTAAA	702
DB	837	CAAGATTTGGCTTTCTCCTGAGTGGATTTCCCAACATTCAGGAAACAAATCAACTAAA	896
QY	703	ACAAATTTGCCAGAAATCTTTTCTCTAAAGTCAAACTGTTGCAGTTGATATTCAGAGCC	762
DB	897	ACAAATTTGCCAGAAATCTTTTCTCTAAAGTCAAACTGTTGCAGTTGATATTCAGAGCC	956
QY	763	AGTGAGGCTTCTTAAACTGAATAACCCCAAGGATTTTCAAGATTTGAATTAACCAACT	822
DB	957	AGTGAGGCTTCTTAAACTGAATAACCCCAAGGATTTTCAAGATTTGAATTAACCAACT	1016
QY	823	AAGAAGACATGACCATTTGATGGAAGAACTGACCATTAAGTCTCTGCATATTTATTATGG	882
DB	1017	AAGAAGACATGACCATTTGATGGAAGAACTGACCATTAAGTCTCTGCATATTTATTATGG	1076

QY	883	GATCTGAGCGCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGA	942
DB	1077	GATCTGAGCGCATCAGCCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCGTTTGA	1136
QY	943	GATACGAGAACTAGGTTACTCATTTGGATCTATCGAGAGCATGACCAATGCGTTCGG	1002
DB	1137	GATACGAGAACTAGGTTACTCATTTGGATCTATCGAGAGCATGACCAATGCGTTCGG	1196
QY	1003	TAATATTTTCAATGTCACCAAGGGCAGATTCATCTCCCTGGCTGAACCTTTGACAACTCC	1062
DB	1197	TAATATTTTCAATGTCACCAAGGGCAGATTCATCTCCCTGGCTGAACCTTTGACAACTCC	1256
QY	1063	GTGACAATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTTACCTTAGT	1122
DB	1257	GTGACAATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCCTTACCTTAGT	1316
QY	1123	TCACCTGCTATTTGAAGTCACTTCATCGCATCGAAGGCTGTCCAGAAAGTTTATTATAC	1182
DB	1317	TCACCTGCTATTTGAAGTCACTTCATCGCATCGAAGGCTGTCCAGAAAGTTTATTATAC	1376
QY	1183	CTAATGATGATGTCATGTTTGGGAAGGATGCTCGGCCAGATGATTTTACAGTCACTCC	1242
DB	1377	CTAATGATGATGTCATGTTTGGGAAGGATGCTCGGCCAGATGATTTTACAGTCACTCC	1436
QY	1243	AAAGCCAGAAAGGTTTATTTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGT	1302
DB	1437	AAAGCCAGAAAGGTTTATTTGACATGCGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGT	1496
QY	1303	TCCTGGGATTAAGGATGCTTATTTGACAAGGCTTTGTAATTAATTCAGCCTCGGATGGGAT	1362
DB	1497	TCCTGGGATTAAGGATGCTTATTTGACAAGGCTTTGTAATTAATTCAGCCTCGGATGGGAT	1556
QY	1363	GGTGGGATTTGCTCTGGAACAGTGAGGAGTGCCTATATTTGACGAGGTGAGGAGTACT	1422
DB	1557	GGTGGGATTTGCTCTGGAACAGTGAGGAGTGCCTATATTTGACGAGGTGAGGAGTACT	1616
QY	1423	GGGAGTATTTGGAGTTGACAGCCCTGCGAGTTTGGTGGAGAAATAAACAAGTCTCTTAC	1482
DB	1617	GGGAGTATTTGGAGTTGACAGCCCTGCGAGTTTGGTGGAGAAATAAACAAGTCTCTTAC	1676
QY	1483	TGTAATCAGGATGTCGGAATTTCTCGCTCGCTGATAGTTCTGTGACCAAGCATGCAAT	1542
DB	1677	TGTAATCAGGATGTCGGAATTTCTCGCTCGCTGATAGTTCTGTGACCAAGCATGCAAT	1736
QY	1543	GTCCTGCTGCTGGGTTTGTGCTGCGACTGTGGCAAGATCATTTTCAATGATGAT	1602
DB	1737	GTCCTGCTGCTGGGTTTGTGCTGCGACTGTGGCAAGATCATTTTCAATGATGAT	1796
QY	1603	AAAGTGAATCTTCTCCCAACACAGACTCATATATTTATCCAAAGGTGAATGCCCT	1662
DB	1797	AAAGTGAATCTTCTCCCAACACAGACTCATATATTTATCCAAAGGTGAATGCCCT	1856
QY	1663	TATTTACAGCTTTGAGAAAGTAGCAAAAGAGGAGTTGAAGGTGCCTATAGTGACAAATCCA	1722
DB	1857	TATTTACAGCTTTGAGAAAGTAGCAAAAGAGGAGTTGAAGGTGCCTATAGTGACAAATCCA	1916
QY	1723	ATTAATTCGACATGCTTCTATTTGCCAACAGTGGAAACCACTCCATCATATATGACAGT	1782
DB	1917	ATTAATTCGACATGCTTCTATTTGCCAACAGTGGAAACCACTCCATCATATATGACAGT	1976
QY	1783	GGATGAATGCCACCAATACATTTTATCTCACCTTTCAAAATACAAACATGAAGAG	1842
DB	1977	GGATGAATGCCACCAATACATTTTATCTCACCTTTCAAAATACAAACATGAAGAG	2036
QY	1843	TTCAAAATGCAGATAACAGTGGAGGTGGACAAAGGAGGACCAAACTGAATTTCTAGC	1902
DB	2037	TTCAAAATGCAGATAACAGTGGAGGTGGACAAAGGAGGACCAAACTGAATTTCTAGC	2096
QY	1903	GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGCGGAAATC	1962
DB	2097	GCCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGCGGAAATC	2156

Qy	1963	CTTTTGGAGATATCCCAAAGAAAACGCTTCCCGAAGTTTAAAGACATGATGTTAAC	2022
Db	2157	CTTTTGGAGATATCCCAAAGAAAAGCGCTTCCCGAAGTTTAAAGACATGATGTTAAC	2216
Qy	2023	TCAACAAGGAGAGCCCGAGGAAGAGGTGAAAATTCCTCGTGGTAAATATTTTCACCTCTTCCA	2082
Db	2217	TCAACAAGGAGAGCCCGAGGAAGAGGTGAAAATTCCTCGTGGTAAATATTTTCACCTCTTCCA	2276
Qy	2083	AAAGACGCCACGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACT	2142
Db	2277	AAAGACGCCACGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACTGGAGACATCACT	2336
Qy	2143	TTGAAGGATACAAATTTGTCCAAAGTCAGCCTTGCTGAGATCATTTCTGATCAACTCACAG	2202
Db	2337	TTGAAGGATACAAATTTGTCCAAAGTCAGCCTTGCTGAGATCATTTCTGATCAACTCACAG	2396
Qy	2203	CATGCTAAAAATAAAATCAAGCTATAATAACAGATGAACAAATAGACGTTGGTGGCT	2262
Db	2397	CATGCTAAAAATAAAATCAAGCTATAATAACAGATGAACAAATAGACGTTGGTGGCT	2456
Qy	2263	CCACAGAAAACAGGTTCATAAAGCATCTTGCCAAACAGCTTGGAGTGTCTGGAAGA	2322
Db	2457	CCACAGAAAACAGGTTCATAAAGCATCTTGCCAAACAGCTTGGAGTGTCTGGAAGA	2516
Qy	2323	TTGCAGAGGTTGACTTTTCCCTGCAGCTGAGTGTAAAGTGAATGTGTCATGCACAGGGTCAG	2382
Db	2517	TTGCAGAGGTTGACTTTTCCCTGCAGCTGAGTGTAAAGTGAATGTGTCATGCACAGGGTCAG	2576
Qy	2383	AATCCACCCCTGGACCTTGAGACACACAGATTTAGAGTGGAACTCACACCCCAAAA	2636
Db	2577	AATCCACCCCTGGACCTTGAGACACACAGATTTAGAGTGGAACTCACACCCCAAAA	2696
Qy	2443	ACCATAGGCGGAATCTGCACAAAGAAAAGCCCCCATCTCTGATTGTTCACCTGGAAGC	2502
Db	2637	ACCATAGGCGGAATCTGCACAAAGAAAAGCCCCCATCTCTGATTGTTCACCTGGAAGC	2696
Qy	2503	CAGATGACAAAAAGAAAAGAAAATCACAGGGAAAGAAAAGAGACAGTAGAATGGAGAA	2562
Db	2697	CAGATGACAAAAAGAAAAGAAAATCACAGGGAAAGAAAAGAGACAGTAGAATGGAGAA	2756
Qy	2563	AATGCTCAAAATCACATAGCGCTTACTGAAGTGTTACTTTGGAAGAAAGCTGCACGATTAC	2622
Db	2757	AATGCTCAAAATCACATAGCGCTTACTGAAGTGTTACTTTGGAAGAAAGCTGCACGATTAC	2816
Qy	2623	ACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAGTATTTCTAGATCTTCTC	2682
Db	2817	ACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAAAGTATTTCTCAAGATCTTCTC	2876
Qy	2683	GAGAGAAGAGTCATTTGAAGACACAATTTGGCCTACTTCTACTGATGCAAGATAGAGCC	2742
Db	2877	GAGAGAAGAGTCATTTGAAGACACAATTTGGCCTACTTCTACTGATGCAAGATAGAGCC	2936
Qy	2743	AGATAACAAGAGATACATTTGCAGATTCCTTCAGATGTAAATATAAATTTCTAAATAGC	2802
Db	2937	AGGCACTAAAGATACATTTGCAGATTCCTTCAGATGTAAATATAAATTTCTAAATAGC	2996
Qy	2803	AAGTTTGGATTCACATCGCGGAAGATCCCTGCTCAGATGCCTCACAGATGATGACCGGATT	2862
Db	2997	AAGTTTGGATTCACATCGCGGAAGATCCCTGCTCAGATGCCTCACATGATGATGACCGGATT	3056
Qy	2863	GTTATGCAAGACTGCAAGATATGTTCCCTGAAGAAATTTGACAGAGCTGATTTACAAA	2922
Db	3057	GTTATGCAAGACTGCAAGATATGTTCCCTGAAGAAATTTGACAGAGCTGATTTACAAA	3116
Qy	2923	GTGCGCATTTCTGAGGATATGCAAGTTTGCTTCTCTTATTTTATTTCTCATGAGTGCA	2982
Db	3117	GTGCGCATTTCTGAGGATATGCAAGTTTGCTTCTCTTATTTTATTTCTCATGAGTGCA	3176
Qy	2983	GTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTTGATACAGATCAATCTGGTGTC	3042
Db	3177	GTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTTGATACAGATCAATCTGGTGTC	3236
Qy	3043	TTGTCTGACAGAAAATCCGAACACTGGCTACCGAATTTACGAACTCCGCTTAAGTTTG	3102

Db	3237	TTGTCTCAGCAGAGAAATCCGACACCTGGCTCCAGAAATCAGAACTCCGCTTAAGTTG	3296
Qy	3103	CAGGATTTGACAGTCTGGAAACACATGCTAAATAAATGCTCAAAAATGCTTCTGCTGAT	3162
Db	3297	CAGGATTTGACAGTCTGGAAACACATGCTAAATAAATGCTCAAAAATGCTTCTGCTGAT	3356
Qy	3163	ATCAGCAGCTAAATAATATATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA	3222
Db	3357	ATCAGCAGCTAAATAATATATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA	3416
Qy	3223	CCGCTCACTAAAAGTCTAGTAACAAAACTGTAACCAAGTAACCTGACAAAATCCACAAAGCA	3282
Db	3417	CCGCTCACTAAAAGTCTAGTAACAAAACTGTAACCAAGTAACCTGACAAAATCCACAAAGCA	3476
Qy	3283	TATTAAGGACAAAAACAATATATAGTTTGAATCATATGGGAGAGAAGAAATCGCTTTAAA	3342
Db	3477	TATTAAGGACAAAAACAATATATAGTTTGAATCATATGGGAGAGAAGAAATCGCTTTAAA	3536
Qy	3343	ATGATTCGTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAACCCCT	3402
Db	3537	ATGATTCGTACCAACGTTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAACCCCT	3596
Qy	3403	AGGAAGTTGTTGGCTCGAATGACAACATTCAGCACATCATAAAGATGCTCAGACAGTG	3462
Db	3597	AGGAAGTTGTTGGCTCGAATGACAACATTCAGCACATCATAAAGATGCTCAGACAGTG	3656
Qy	3463	AAGGCTCTTCAGGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG	3522
Db	3657	AAGGCTCTTCAGGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG	3716
Qy	3523	CCAAGAGAGTATCGAAACCGTTTTCTTCATATGATGAGCTCAGGAATGGAGGCTTAT	3582
Db	3717	CCAAGAGAGTATCGAAACCGTTTTCTTCATATGATGAGCTCAGGAATGGAGGCTTAT	3776
Qy	3583	CGAGACAAATTGAAGT	3598
Db	3777	CGAGACAAATTGAAGT	3792
RESULT 6			
US-09-986-552-4			
; Sequence 4, Application US/09986552			
; Patent No. US20020150981A1			
; GENERAL INFORMATION:			
; APPLICANT: CANFIELD, William			
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL			
; FILE REFERENCE: 215089US77DIV			
; CURRENT APPLICATION NUMBER: US/09/986,552			
; CURRENT FILING DATE: 2001-11-09			
; PRIOR APPLICATION NUMBER: 09/635,872			
; PRIOR FILING DATE: 2000-08-10			
; PRIOR APPLICATION NUMBER: 60/153,831			
; PRIOR FILING DATE: 1999-09-14			
; NUMBER OF SEQ ID NOS: 52			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 4			
; LENGTH: 5597			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-986-552-4			

	Query Match	96.3%	Score 3465.6	DB 10	Length 5597
	Best Local Similarity	99.5%	Pred. No. 0		
	Matches 3477	Conservative	0	Mismatches 19	Indels 0
				Gaps	
QY	103	ACCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAAATATTCCTGGAAGTCC	162		
Db	297	AGCCGAGATCAATACCATGTTTGTGTTGATTCCTATAGAGACAAATATTCCTGGAAGTCC	356		
QY	163	TTTCAGATCGGCTTTGTCGCCATGCCGATGACGTGTTTACACCTGGGTGAATGSC	222		
Db	357	TTTCAGATCGGCTTTGTCGCCATGCCGATGACGTGTTTACACCTGGGTGAATGSC	416		



QY 223 ACAGATCTTGAACCTACTCAAGGAACCTACAGCAGGTGACAGAACAGATGGAGGAGGCAG 282  
DB 417 ACAGATCTTGAACCTACTCAAGGAACCTACAGCAGGTGACAGAACAGATGGAGGAGGCAG 476  
QY 283 AAAGCAATGAGAGAAATCCTTTGGGAAAAACACAAACGGAACCTACTAAGAAGAGTGAGAAG 342  
DB 477 AAAGCAATGAGAGAAATCCTTTGGGAAAAACACAAACGGAACCTACTAAGAAGAGTGAGAAG 536  
QY 343 CAGTTAGAGTGTTCCTTAACACACTGCATTAAAGTGCCATGCTTGTCTACTGCGCAGGCC 402  
DB 537 CAGTTAGAGTGTTCCTTAACACACTGCATTAAAGTGCCATGCTTGTCTACTGCGCAGGCC 596  
QY 403 CTGCCAGCCAACTACCCCTGAAGGAGCTGCCATCTCTTTATCCTCTTTTCAATCTGTC 462  
DB 597 CTGCCAGCCAACTACCCCTGAAGGAGCTGCCATCTCTTTATCCTCTTTTCAATCTGTC 656  
QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 522  
DB 657 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTTGTTGTT 716  
QY 523 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGA 582  
DB 717 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTGGACTGCTTAAAGGAAATAGCAGA 776  
QY 583 CAGACAGTATGGAGGGCTACTTTGACAAACAGATAAAGAGTCCCTGGATTTAGTCTAATG 642  
DB 777 CAGACAGTATGGAGGGGTACTTTGACAAACAGATAAAGAGTCCCTGGATTTAGTCTAATG 836  
QY 643 CAAGATTTGGCTTTCTGAGTGGATTTCACCAACATTTCAAGGAAACAAATCAACTAAAA 702  
DB 837 CAAGATTTGGCTTTCTGAGTGGATTTCACCAACATTTCAAGGAAACAAATCAACTAAAA 896  
QY 703 ACAAATTTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 762  
DB 897 ACAAATTTGCCAGAAATCTTTCTCTAAAGTCAAACTGTTGAGTGTATTCAGAGGCC 956  
QY 763 AGTGTAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 822  
DB 957 AGTGTAGCGCTTCTAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 1016  
QY 823 AAGAGAACATGACCATTTGATGGAAGAAGACTGACCATAGTCCCTGCATTTATTTATGG 882  
DB 1017 AAGAGAACATGACCATTTGATGGAAGAAGACTGACCATAGTCCCTGCATTTATTTATGG 1076  
QY 883 GATCTGAGCCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTCCAGTCTGTTTGA 942  
DB 1077 GATCTGAGCCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTCCAGTCTGTTTGA 1136  
QY 943 GATAACGAAGACTGAGGTACTCATTTGCGATCTATCGAGGCGATGCACCATGGGTTCCG 1002  
DB 1137 GATAACGAAGACTGAGGTACTCATTTGCGATCTATCGAGGCGATGCACCATGGGTTCCG 1196  
QY 1003 AATATTTTCATTTGTCACCAACGGGAGATTCATCTCGTGAACCTTGACAACTCCGA 1062  
DB 1197 AATATTTTCATTTGTCACCAACGGGAGATTCATCTCGTGAACCTTGACAACTCCGA 1256  
QY 1063 GTGACATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACCTGCTACCTTTAGT 1122  
DB 1257 GTGACATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACCTGCTACCTTTAGT 1316  
QY 1123 TCACCTGCTATGAAAGTACAGTTCAATCGCATCGAAGGCGTGTCCAGAAAGTTTATTTAC 1182  
DB 1317 TCACCTGCTATGAAAGTACAGTTCAATCGCATCGAAGGCGTGTCCAGAAAGTTTATTTAC 1376  
QY 1183 CTAATATGATCATGTCATGTTTGGGAAGGATGCTGCCAGATGATTTTACAGTCACTCC 1242  
DB 1377 CTAATATGATCATGTCATGTTTGGGAAGGATGCTGCCAGATGATTTTACAGTCACTCC 1436  
QY 1243 AAAGGCCAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGCGTCCAGAGT 1302  
DB 1437 AAAGGCCAGAGGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGCGTCCAGAGT 1496

QY 1303 TCCTGATTTAAGGATGGCTATTGTTGACAAAGGCTTGTAAATAATTCAGCCTGCGATTTGGAT 1362  
DB 1497 TCCTGATTTAAGGATGGCTATTGTTGACAAAGGCTTGTAAATAATTCAGCCTGCGATTTGGAT 1556  
QY 1363 GGTGGGATTTGCTCTGGAACACAGTGGAGGACTCCCTATATTGCAAGGAGTGGAGGTACT 1422  
DB 1557 GGTGGGATTTGCTCTGGAACACAGTGGAGGACTCCCTATATTGCAAGGAGTGGAGGTACT 1616  
QY 1423 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTCTCTCTTAC 1482  
DB 1617 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTCTCTCTTAC 1676  
QY 1483 TGTAAATCAGGATGTCGGAATTTCCCTGGCTGCTGATAAGTCTGTGACCAAGCATGCAAT 1542  
DB 1677 TGTAAATCAGGATGTCGGAATTTCCCTGGCTGCTGATAAGTCTGTGACCAAGCATGCAAT 1736  
QY 1543 GTCTTGTCTGTTGGTTTGATGCTGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT 1602  
DB 1737 GTCTTGTCTGTTGGTTTGATGCTGCGACTGTGGCAAGATCAATTTTCATGAATTTGAT 1796  
QY 1603 AAAGTATGCTCTTCTCCAAACAGACTCAGTATATTTATTTCCAAAAGTGAATGCTGCT 1662  
DB 1797 AAAGTATGCTCTTCTCCAAACAGACTCAGTATATTTATTTCCAAAAGTGAATGCTGCT 1856  
QY 1663 TATTTTCAGCTTTGCAAGAGTAGCCAAAGAGAGGAGTTGAAGTGCCTATAGTGACAAATCCA 1722  
DB 1857 TATTTTCAGCTTTGCAAGAGTAGCCAAAGAGAGGAGTTGAAGTGCCTATAGTGACAAATCCA 1916  
QY 1723 ATAAATTCGACATGCTTCTATTGCCAACAAAGTGGAAACCATCCACTCATATGACACAGT 1782  
DB 1917 ATAAATTCGACATGCTTCTATTGCCAACAAAGTGGAAACCATCCACTCATATGACACAGT 1976  
QY 1783 GGAATGAATGCCACACAAATACATTTTAACTCTACGTTTCAAAAATACAAACGATGAAGAG 1842  
DB 1977 GGAATGAATGCCACACAAATACATTTTAACTCTACGTTTCAAAAATACAAACGATGAAGAG 2036  
QY 1843 TTCAAAATGACAGATAACAGTGGAGGTGGACACAAAGGGAGGACCAAACTGAAATCTACG 1902  
DB 2037 TTCAAAATGACAGATAACAGTGGAGGTGGACACAAAGGGAGGACCAAACTGAAATCTACG 2096  
QY 1903 GCCCAGAGGGTTAGCAAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGCGGAAATC 1962  
DB 2097 GCCCAGAGGGTTAGCAAAATTTAGTTAGTCCCATTAACACTTCTTCCAGAGCGGAAATC 2156  
QY 1963 CTTTITTTAGGATATTTCCCAAGAAACAGCTTCCGGAAGTTTAAGAGACATGATGTTAAC 2022  
DB 2157 CTTTITTTAGGATATTTCCCAAGAAACAGCTTCCGGAAGTTTAAGAGACATGATGTTAAC 2216  
QY 2023 TCAACAAGGAGAGCCAGGAAGGTGAAATTTCCCTGGTAAATATTTCACTCTCTTCCA 2082  
DB 2217 TCAACAAGGAGAGCCAGGAAGGTGAAATTTCCCTGGTAAATATTTCACTCTCTTCCA 2276  
QY 2083 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2142  
DB 2277 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2336  
QY 2143 TTGAAGAGATACAAATTTTCCAAAGTCACTTGTGAGATCATTTCTGATGAATCAACAG 2202  
DB 2337 TTGAAGAGATACAAATTTTCCAAAGTCACTTGTGAGATCATTTCTGATGAATCAACAG 2396  
QY 2203 CATGCTAAATATAAAATCAAGCTTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT 2262  
DB 2397 CATGCTAAATATAAAATCAAGCTTATAATAACAGATGAAACAAATGACAGTTTGGTGGCT 2456  
QY 2263 CCACAGGAAAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAGA 2322  
DB 2457 CCACAGGAAAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAGA 2516  
QY 2323 TTGCAGAGGTTGACTTTTCTTCCCTGCAGTGTAAAGTGAATGTCATGACCGAGGTCAG 2382  
DB 2517 TTGCAGAGGTTGACTTTTCTTCCCTGCAGTGTAAAGTGAATGTCATGACCGAGGTCAG 2576  
QY 2383 AATCCACCCCTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAAACTCAGACCCCAAAA 2442



Db 2577 AATCCACCCTGGACTGGAGACCACAGCAAGATTAGAGTGAACCTCACACCACAAAA 2636  
QY 2443 ACCATAGCGGAAATGTGACAAAGAAAAAGCCCCCATCTCTGATGTGTTCCACTGGAAGC 2502  
Db 2637 ACCATAGCGGAAATGTGACAAAGAAAAAGCCCCCATCTCTGATGTGTTCCACTGGAAGC 2696  
QY 2503 CAGATGACAAAAAGAAAAATTCACAGGGAAGAAAAAGAGAACAGTAGAATGGAGAA 2562  
Db 2697 CAGATGACAAAAAGAAAAATTCACAGGGAAGAAAAAGAGAACAGTAGAATGGAGAA 2756  
QY 2563 AATGCTGAAATACATAGCGCTTACTGAAGTGTACTTTGGAAGAAAGCTCGAGCAATPAC 2622  
Db 2757 AATGCTGAAATACATAGCGCTTACTGAAGTGTACTTTGGAAGAAAGCTCGAGCAATPAC 2816  
QY 2623 ACAGATAGTACTTGGCTTTTGGCCATGGGAGAAAAAGTATTTCTAGATCTTCTC 2682  
Db 2817 ACAGATAGTACTTGGCTTTTGGCCATGGGAGAAAAAGTATTTCTAGATCTTCTC 2876  
QY 2683 GACGAAGAGAGTCAATGAAGACACAATTTGGCTTACTTCACTGATAGCAAGAAATAGAGCC 2742  
Db 2877 GACGAAGAGAGTCAATGAAGACACAATTTGGCATACTTCACTGATAGCAAAATACTGGG 2936  
QY 2743 AGATACAGAGAGATACATTTCCAGATTCCCTCAGATATGTAATAAATTTCTAATAGC 2802  
Db 2937 AGGCAACTAAAAGATACATTTCCAGATTCCCTCAGATATGTAATAAATTTCTAATAGC 2996  
QY 2803 AAGTTTGGATTACATCGCGGAATCCCTGCTACATGCTCAGATGATGACCGGATT 2862  
Db 2997 AAGTTTGGATTACATCGCGGAATCCCTGCTACATGCTCAGATGATGACCGGATT 3056  
QY 2863 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGACGCTCATTTTACAAA 2922  
Db 3057 GTTATGCAAGAACTGCAAGATATGTTCCCTGAAGAAATTTGACAAGACGCTCATTTTACAAA 3116  
QY 2923 GTGCGCCATTCGAGATATGCAAGTTGGCTTCTCTTATTTTATATCTCANGAGTGA 2982  
Db 3117 GTGCGCCATTCGAGATATGCAAGTTGGCTTCTCTTATTTTATATCTCANGAGTGA 3176  
QY 2983 GTGACGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTC 3042  
Db 3177 GTGACGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTC 3236  
QY 3043 TTGTCTGACAGAGAAATCCGAACACTGGCTACCAGAAATTCACGAATTCACGCTTAAGTTTG 3102  
Db 3237 TTGTCTGACAGAGAAATCCGAACACTGGCTACCAGAAATTCACGAATTCACGCTTAAGTTTG 3296  
QY 3103 CAGGATTTGACAGGCTGGAACACATGCTAATAATTTGCTCAAAAATGCTTCTCTGCTGAT 3162  
Db 3297 CAGGATTTGACAGGCTGGAACACATGCTAATAATTTGCTCAAAAATGCTTCTCTGCTGAT 3356  
QY 3163 ATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA 3222  
Db 3357 ATCAGCAGCTAAATAATATTCACCAACTCAGGAATCCTACTATGATCCCAACCTGCCA 3416  
QY 3223 CCGGTCACTAAAGTCTAGTAAACAACCTGTAACCAAGTAAGTACGAAAAATCCACAAAGCA 3282  
Db 3417 CCGGTCACTAAAGTCTAGTAAACAACCTGTAACCAAGTAAGTACGAAAAATCCACAAAGCA 3476  
QY 3283 TATAAGGACAAAAACAATATAGTTTGAATTCATGGGAGAGAAATCGCTTTTAAA 3342  
Db 3477 TATAAGGACAAAAACAATATAGTTTGAATTCATGGGAGAGAAATCGCTTTTAAA 3536  
QY 3343 ATGATTCGTACCAAGCTTTCTCATGTGTTGGCCAGTTGGATGACATGAAGAAAAAACCT 3402  
Db 3537 ATGATTCGTACCAAGCTTTCTCATGTGTTGGCCAGTTGGATGACATGAAGAAAAAACCT 3596  
QY 3403 AGGAAGTTGTTGGCTGAATGACAAACATTTGACCAACATATCAATGATGCTCAGACAGTG 3462  
Db 3597 AGGAAGTTGTTGGCTGAATGACAAACATTTGACCAACATATCAATGATGCTCAGACAGTG 3656  
QY 3463 AAGGCTGTCTCAGGACTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTG 3522

Db 3657 AAGCTCTTCTCAGGGACTTCTATGAATCATGTTCCCATACCTTCCCAATTTGAACGTG 3716  
QY 3523 CCAAGAGAGTATCGAAACCGTTTCTTTCATATGATGATGATGATGATGATGATGATGAT 3582  
Db 3717 CCAAGAGAGTATCGAAACCGTTTCTTTCATATGATGATGATGATGATGATGATGAT 3776  
QY 3583 CGAGACAAATTTGAAGT 3598  
Db 3777 CGAGACAAATTTGAAGT 3792

## RESULT 7

US-09-895-072-16  
; Sequence 16, Application US/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLAS  
; FILE REFERENCE: 210119USOCONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 16  
; LENGTH: 5229  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-895-072-16

Query Match 65.6%; Score 2362.6; DB 10; Length 5229;  
Best Local Similarity 81.2%; Pred. No. 0;  
Matches 2842; Conservative 1; Mismatches 590; Indels 66; Gaps 6;

QY 103 AGCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGACACAATATTTGCGTGAAGTCC 162  
Db 148 AGCGAGATCAGTACCATGTTTGTGTTTGTATTCCTATAGACACAATATTTGCGTGAAGTCC 207  
QY 163 TTTGAGATCGGCTTTGCTGCCCCATGCCGATGAGCTGTTTACACCTGGGTGAATGGC 222  
Db 208 TTTGAGATCGGCTTTGCTGCCCCATGCCGATGAGCTGTTTACACCTGGGTGAATGGC 267  
QY 223 ACAGATCTTGAACCTACTGGAAGAACTACAGAGGTACAGAGACAGATGGAAGAGAGCAG 282  
Db 268 ACTGACCTTGAACCTACTGGAAGAACTACAGAGGTACAGAGACAGATGGAAGAGAGCAG 327  
QY 283 AAAGCAATGAGAAATTCCTTTGGGAAAAACACAAACGGAACCTACTAAGAGAGTGAAG 342  
Db 328 AGAGCCATCGGGAACCCCTCGGGAAGAACACAAACGGAACCTACTAAGAGAGTGAAG 387  
QY 343 CAGTTAGAGTGTGCTTAACACACATGCAATTAAGTGGCCATGCTTCTGACCCAGCC 402  
Db 388 CAGTTAGAGTGTGCTGTCGACGACATGCAATTAAGTGGCCATGCTTCTGACCCAGCC 447  
QY 403 CTGCGACCCCAACATCACCTCGAAGGACCTGCCATCTCTTTATCTCTTCTTCTTCTGCG 462  
Db 448 CTGCGACCCCAACATCACCTCGAAGGACCTGCCATCTCTTCTTCTTCTTCTTCTGCG 507  
QY 463 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTGTGTT 522  
Db 508 AGCGACATGTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTGTGTT 567  
QY 523 TTTGACAGTACTAAGGATGTTGAAGATGCCCCTCTGGAAGTCTTAAAGGAAATAGCAG 582  
Db 568 TTTGACAGTACTAAGGATGTTGAAGATGCCCCTCTGGAAGTCTTAAAGGAAATAGCAG 627  
QY 583 CAGACAGTATGAGGGGCTACTTGTACAAACAGATAAAGAGTCCCTGATGATGCTAATG 642  
Db 628 ACAGATGTTTGGAGAGCTTACTTGTACAAACAGATAAAGAGTCCCTGATGATGCTAATG 687

QY	643	CAAGATTGGCTTCTCTGAGTGGATTTCACCAACAATTC	CAAGAAACAATCAACTAAAA	703
Db	688	CAAGCTTTGGGGTTCTCTGAGTGGATTCCACCGACCTT	CTCAAGSAGACGAGTCAACTGAAG	747
		CAAAAATGGCCAGAAAATCTTTCCTCT	--AAAGTCAAACCTGTTGCAGTGTGATTTCAGAG	759
Db	748	ACAAGCTGCCAGAAAAGCTTTCCCTCTAAAAATAAGCT	GTGCGGTGTACTTCGGAG	807
QY	760	GCCAGTGTAGCGCTTCTTAAACCTGAATAACCCCAAGG	ATTTTCAAGAATTGAATAAGCAA	819
Db	808	GCCAGTGTGCTCTTCTGAAATTGAATATCCCAAGGG	TTTCCAGAGACTGAACAAGCG	867
QY	820	ACTAAGAARACATGACCATTTGATGGAAGAARACT	GACCAATAGTCTGATTTATTTATTA	879
Db	868	ACCAAGAAGACATGACCATGATGGAGAGAACTGACCA	TGCGCTGAGCCCTGCGTATCTGCTG	927
QY	880	TGGATCTGAGCGCCATCAGCCAGTCTAAGCAGGAT	GAAGACATCTCTGCCAGTGGTTTT	939
Db	928	TGGGACTGAGTGGCATCGCCAGTCCAGCAGGAT	GAGGAGCGCTCTGCCAGCGCTTT	987
QY	940	GAGATACAAGAAACTGAGGTACTCATTTGCGATCT	TATCGAGAGCATGCACATGGGTT	999
Db	988	GAGGATAATGAAGAGCTGAGGTACTCGCTCGCAT	CTATCGAGAGACACGCGCATGGGA	1047
QY	1000	CGGATATTTTCATTTGTCACCAAGGGGAGATTC	CAATCGCTGAACTTGACAATCCT	1059
Db	1048	CGGAATATTTTCATTTGTCACCAAGGGGAGATTC	CAATCGCTGAACTTGACAATCCT	1107
QY	1060	CGAGTGAACAATAGTACACACACAGGATGTTTT	TGCAAAATTTGAGCCACTTGCCCTACCTTT	1119
Db	1108	CGAGTGACCATAGTACCCACAGGACATTTTCC	AAAAATCTGAGGCATTTGCCCTACTTTC	1167
QY	1120	AGTTACCTGCTATTGAAAGTCAAGTTCATCGAT	CTCGATCGAAGGCTGTCCCAAGTTTATT	1179
Db	1168	AGTTCCCTGCTATTGAAAGTCAATTCACCGCAT	CCAGGCTGTCCCAAGTTTATT	1227
QY	1180	TACTTAATGATGATGATGTTTGGGAAGGATGCT	TGGCCAGATGATTTTTACAGTCAAC	1239
Db	1228	TATCTAAATGACCATGATGTTTCGGTAAGSAGCT	CTGCGCGGAGGATTTTACAGCCAC	1287
QY	1240	TCCAAAGGCCAAGGTTTATTTGACATGGCTGT	GCCAACTGTGCCAGGGCTGCCCA	1299
Db	1288	TCCAAAGGTCAAAAGTTTATTTGACATGGCTGT	GCCAACTGTGCCAGGGCTGCCCG	1347
QY	1300	GGTTCCTGGATTAAAGATGGCTATTGTGACA	AGGCTTGTAATTCAGCCCTGCGATTGG	1359
Db	1348	GGCTCTGGATTAAAGACGGCTATTGTGATA	AGGCTTGTAATTCAGCCCTGCTACTGG	1407
QY	1360	GATGGTGGGATTGCTCTGGAACAAGTGGAGG	AGTCCGTATATTCAGAGAGTGGAGGT	1419
Db	1408	GATGGGGGAAACTGCTCTGGTAACTACTG	CAGGGAAACCGGTTGTTCAGAGTGGGGGT	1467
QY	1420	ACTGGGAGTATTGGAGTTGGACAGCCCTGG	CAGTGTGGAGGATTAACAGTGTCTCT	1479
Db	1468	ACCGGAAATATGGAGCTGGACAGCAGTGG	CAGTGTGGAGGATTAACAGCATTCTCT	1527
QY	1480	TACTGTAATCAGGATGTGCGAATTCCTGGCT	TCGCTGATTAAGTCTGTGACCAAGCATGC	1539
Db	1528	TACTGTAACCAAGGATGTGCAAACTCCTGG	CTGGCTGATTAAGTCTGTGACCAAGCCTGT	1587
QY	1540	AATGTCTTGTCTGGGTTTGATGCTGCGACT	GTGGCAAGATCAATTTTCATGAATTG	1599
Db	1588	AAGCTCTTATCTGCGGGTTTGATGCTGGT	GACTGTGCACAAGATCAATTTTCATGAATTG	1647
QY	1600	TATAAAGTGTATCTTCTCCCAACAGACTCA	CTATATATTTCAAAAGTGAATGCTGT	1659
Db	1648	TATAAAGTAACTTCTCCCAACAGACTCACT	ATGTTGTCCCAAGTGAATGCTGT	1707
QY	1660	CCTTATTTGAGTTTTCAGAACTAGC	AAAAAGAGGTTGAAGTGGCTATAGTACAAT	1719
Db	1708	TCCTTATTCAGTTTTCAGAACTAGC	AGAAAGATTTGAAGGAGCTACAGCGCAAC	1767
QY	1720	CCAATAATTCGACATGCTTCTATTGCC	AACAAGTGGAAAAACCATCCACCTCAATAATGCAC	1779

[illegible]

Db	2785	AGCAAGTTTGGATTACATCCAGGAAGAGTCCCTGCACACATGCCGCACATGAITTGACAGG	2844
Qy	2860	ATTGTTATGCAAGAACTGCAAGATATGTTCCTGAAGAATTTGACAAGACGCTCATTTTCAC	2919
Db	2845	ATCGTTATGCAAGAAGCTCCAAAGATATGTTCCTGAAGAATTTGACAAGACTTTCATTTTCAC	2904
Qy	2920	AAAGTGGCGCATTTCTGAGGATATGCAAGTTTGGCTTCTTATATTTTATATATCTCATGAGT	2979
Db	2905	AAGGTGCGTCACTCTCAGGACATGCAAGTTTGGCTTCTCCTACTATTTATACCTCATGAGT	2964
Qy	2980	GCAGTGCAGCCCACTGAATATATCTCAAGTCTTTGATGAAGTTGTATACAGATCAATCTGGT	3039
Db	2965	GCAGTTGACGCCCTCAATATTTCCCAAGTCTTTTCATGAAGTAGACACAGACAATCTGGT	3024
Qy	3040	GTCTTTGCTGCACAGAGAAATCCGAACACTGGCTACCGCAATTCACGAATCCGCTTAAAGT	3099
Db	3025	GTCTTTGCTGATAGGAATCCGAACWCTGGCCACGAAATTCAGGACCTTACCTTTAAGC	3084
Qy	3100	TTGCAGGATTTGACAGGTCTGGAACACATGCTTAATAATTTGCTCAAAAAGTCTTCGTGCT	3159
Db	3085	TTGCAGGATTTGACAGGTTTGGAACACATGTTTAATAAATTTGCTCAAAAATGCTCCCGCT	3144
Qy	3160	GATATCACGCAGCTAAATAATATTCCACCACTCAGGAATCCTACTATGATCCCAACCTG	3219
Db	3145	AATATCACTCAACTCAACAACTCCCCCGACTCAGGAAGCATACTACGAGCCCAACCTG	3204
Qy	3220	CCACGGGTCACTAAAAGTCTAGTAAACAAACTGTAACACAGTAAGTGCACAAAATCCACAAA	3279
Db	3205	CCTCCGGTCACTAAGAGTCTTGTCAACCACTGTAGCCAGTAACTGACAAAGATCCACAAA	3264
Qy	3280	GCATATAAGGACAAAACAAATATAGTTTGAATTCATGGGAGAGAGAATAATCGCTTTT	3339
Db	3265	GCCTATAAAGACAAAGACAAATACAGGTTTGAATTCATGGGAGAGAGAATAATCGCTTTT	3324
Qy	3340	AAATGATTCGTACCACGCTTCTCATGTGGTTGGCCAGTTGGATGACATAAGAAAAAC	3399
Db	3325	AAGATGATCGAACCAGTGTCTCATGTGGTTGGTCTAGTTGGATGACATCAGAAAAAC	3384
Qy	3400	CCTAGGAAGTTGTTGGCTGGAATGACAACTGACCACAATCATAAAGATGCTCAGACA	3459
Db	3385	CCCAGGAAGTTCTGTTGCTGGAATGACAACTTGACCACAACATAAAGATGCCGGACA	3444
Qy	3460	GTGAGGCTGTTCTCAGGACATCTATGAATCCATGTTCCCATACCTTCCCAATTTGAA	3519
Db	3445	GTGAAGGCTCTCTCAGGACATCTATGAGTCCATGTTTCCCATACCTTCCAGGTTTGA	3504
Qy	3520	CTGCCAAGAGATGTCGAAACCGTTTCCCTTCATATGTCATGAGCTCAGGAATGGAGGCT	3579
Db	3505	CTGCCAAGAGAGTATCGGAACCGCTTCTGACATGTCATGAGCTCCAGAAATGGCGGCA	3564
Qy	3580	TATCGAGACAAATTTGAAGT	3598
Db	3565	TATCGAGACAAAGCTGAAGT	3583

```

RESULT 8
US-09-986-552-16
; Sequence 16, Application US/09986552
; Patent No. US20020150981A1
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, William
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 215089US77DIV
; CURRENT APPLICATION NUMBER: US/09/986,552
; CURRENT FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: 09/635,872
; PRIOR FILING DATE: 2000-08-10
; PRIOR FILING DATE: 2000-08-10
; PRIOR FILING DATE: 1999-09-14
; PRIOR APPLICATION NUMBER: 60/153,831
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 5229

```

; TYPE: DNA		Query Match		65.6%; Score 2362.6; DB 10; Length 5229;	
; ORGANISM: Mus musculus		Best Local Similarity		81.2%; Pred. No. 0;	
us-09-986-552-16		Matches 284; Conservative		1; Mismatches 590; Indels 66; Gaps	
Qy	103	AGCCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGAGACAATATGCTGGAAGTCC	162		
Db	148	AGCCGAGATCAGTACCATGTTTGTGTTTGTATTCCTACAGACACAATGCTGGGAATCC	207		
Qy	163	TTTCAGAAATCGGCTTGTCTGCCATGCCGATTTGAGTGTGTTTACACCTGGGTGAATGCG	222		
Db	208	TTTCAGAAATCGGCTTGTCTGCCATGCCAATCGAGTGTGTTTACACCTGGGTGAATGCG	267		
Qy	223	ACAGATCTTGAACCTACTGAAGAACTACAGCAGGTGAGAAACAGATGGAGGAGGACG	282		
Db	268	ACTGACCTTGAACCTCTAAAGGAGCTACACGAGTCCGAGAGCACATGGAGGAAGACG	327		
Qy	293	AAAGCAATGAGAGAAATCCTTTGGGAAAAACACAACGGAACCTTACTAAGAGAGTGAAG	342		
Db	328	AGAGCCATGCGGGAACCCCTCGGGAAGAACACAACCGCAACGGAAGAGAGTGAAG	387		
Qy	343	CAGTTAGAGTGTGTTTAAACACACTGCTAATTAAGGTGCCAATGCTGCTTGGAACCCAGCC	402		
Db	398	CAGCTGGATGTCTGCTGAGCCACTGCAATTAAGGTGCCCATGCTGTTCTGGACCCGGCC	447		
Qy	403	CTGCCAGCCAACTACACCTGGAAGGACCTGCCATCTCTTTATCCTTCTTTTCAATCTGCC	462		
Db	448	CTGCCAGCCAACTACACCTGGAAGGATCTGCCAACCCCTTTACCCATCTTTCCACGCGTCC	507		
Qy	463	AGTGACATTTCAATGTTGCAAAACCAAAAACCCCTCTACCAATGCTCAGTTGTGTT	522		
Db	508	AGGACATGTTCAATGTTGCAACCAAAAATCCGCTCTACAATGTCCCGTGTGCTGT	567		
Qy	523	TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGAAATAGCAGA	582		
Db	568	TTTGACACTAGTAAGGATGTTGAAGAGCCCACTGCTGGACGCTTTAAGGAGGCCACGAA	627		
Qy	583	CAGACGATGAGGGGCTACTTGACAACAGATAAAGAGTCCCTCGATTAGTGCTTAATG	642		
Db	628	ACAGATGTTGGAGAGCCTACTTGACAACAGACAAGACGCCCTGGCTTAGTGCTGATA	687		
Qy	643	CAAGATTGGCTTTCCCTGAGTGGATTTCCACCACATTCGAAGAAACAATCAACTAAA	702		
Db	688	CAAGGCTGGCGTTCCCTGAGTGGATTTCCACCAGCCATTCGAAGGAGACGAGTCAACTGAAG	747		
Qy	703	ACAAATTTGCCAGAAATCTTCTCTCT - - AAGTCAAACTGTTCAGTGTGTAATCAGAG	759		
Db	748	ACAAAGTGTGCAAGAAAGTTTCCCTCTAAAAATAAGCTGTTCGGCTGTACTCGGAG	807		
Qy	760	GCCAGTGTAGCGCTCTAAACATGAATTAACCCCAAGGATTTTCAAGAAATTAATAAGCA	819		
Db	808	GCCAGTGTGCTCTTCTGAAATGAATTAATCCCAAGGGTTTCCAAGAGCTGAACAAGCAG	867		
Qy	820	ACTAAGAAGACATGACCATGATGGAAGAAAGACTGACCATAGTCTCGATATTTATTA	879		
Db	868	ACCAAGAGAACATGACCATGATGGAAGAAAGACTGACCATAGTCTCGATATTTATTA	927		
Qy	880	TGGGATCTGAGCGGCATCAGCAGTCTTAAGCAGGATGAAGACATCTCTGCCAGTCTGTT	939		
Db	928	TGGGACCTGAGTGCATCAGCCAGTCCAAGCAGGATGAGACCGCTCTGCCAGCCGTTT	987		
Qy	940	GAGATACGGAAGAACTAGGTAAGTATCGGATCTATCGAGAGGCAATGACCATGGTT	999		
Db	988	GAGGATATGAAGAGCTGAGGTACTCGCTCGATCTATCGAGAGACACGCGCATGGGTA	104		
Qy	1000	CGGAATATTTTCAATTGTCCACACGGGCAATTCATCTCGCTGGCTGAACCTTTGACAACTCT	105		
Db	1048	CGGAATATTTTCAATTGTCCACACGGGCAATTCATCTCGCTGGCTGAACCTTTGACAACTCT	110		
Qy	1060	CGAGTGCAATAGTAACACACCCAGGATGTTTTCGAAATTTGAGCCACTTGCTACTCTTT	111		

Db	1108	CGAGTGCCATGAGTGCACCCACCAGGACATTTTCCAAATCTGAGCCACTTGCTACTTTC	1167
QY	1120	AGTTTCACTGCTATTGAAAGTCACGTTTCATCGCATCGAAGGCGTCTCCCAAGTTTATT	1179
Db	1168	AGTTCCCTGCTATTGAAAGTCACATTCACCGCATCGAAGGCGTCTCCCAAGTTTATT	1227
QY	1180	TACCTAAATGATGATCATGTTTGGGAAGGATGCTGCGCCAGATGATTTTTACAGTCAC	1239
Db	1228	TATCTAAATGACGATGTCATGTTTCGTAAGGACGCTGCGCGGACGATTTTTACAGCCAC	1287
QY	1240	TCCAAAGCCAGAGGTTTATTTGACATGCGCTGTGCCAACTGTGCCAGGCGTGGCCCA	1299
Db	1288	TCCAAAGCTCAAAGGTTTATTTGACATGGCTGTGCCAACTGTGCGAGAGGCGTGGCCG	1347
QY	1300	GGTTCTCGGATTAAAGGATGCTATTGTGCAAGGCTTGTAATAATTCAGCCTGCGATTGG	1359
Db	1348	GGCTCTCTGGATAAAGGACGCGTATTGTGATAAGGCTGTAAATACCTCACCCCTGTGACTGG	1407
QY	1360	GATGGTGGGATTTGCTCTGGAACAGTGGAGGAGTGCCTATATTTGACGAGGTTGAGGT	1419
Db	1408	GATGGCGGAACCTGCTCTGGTATATCTGACGGGAACCGGTTGTGTCAAGAGTGGGGGT	1467
QY	1420	ACTGGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGAGGAATAAACAGTGTCTCT	1479
Db	1468	ACCGGGAATATTGGAGCTGGACAGCACTGGCAGTTTGGTGAGGAATAAACACCATCTCT	1527
QY	1480	TACTGTAAATCAGGATGTGCGAATTCCTGGCTCGCTGATAAAGTTCTGTGACCAAGCATGC	1539
Db	1528	TACTGTAAACCAAGGATGTCAAACTCCTGGCTGGCTGACAGTTCTGTGACCAAGCCTGT	1587
QY	1540	AATGTCTTTGCTGTGGGTTTGANGTGGCGACTGTGGCAAGATCAITTTTCATGAATTG	1599
Db	1588	AACGCTCTATCTCGCGGGTTTATGCTGTGCTACTGTGACAAAGATCATTTTCATGAATTG	1647
QY	1600	TATAAGTGTATCTCTCCCAACACGACTCTACTATATTTTCCAAAGGTTGAATGCGTG	1659
Db	1648	TATAAGTAACTCTCTCCCAACACGACTCTACTATGTTGTCCCAAGGTTGAATGACTGT	1707
QY	1660	CCTTATTTCAGCTTTCAGAACTAGCCAAAGAGAGTTGAAAGTGCTATAGTGACAAT	1719
Db	1708	TCCTTATTTCAGCTTCCAAACATAGCCGAANAAGATTTGAAGGACCTACAGCGACAC	1767
QY	1720	CCAAATATTGCAGATGCTTCTATGCGCAACAGTGGAAAACCATCCACCTCATTAATGCAC	1779
Db	1768	CCCATATCGCGCACGCTCCATTTGCAACAAGTGGAAAACCCCTACACCTGATAATGCC	1827
QY	1780	AGTGGAAATGAATGCCACCACTACATTTTAACTCTCAGTTTCAAATACAAAGCATGAA	1839
Db	1828	GGGGGATGAACGCCACCACTATTTTAACTCTACTCTTCAAAACGCCAACGACGAA	1887
QY	1840	GAGTTCAAATGCAGATAACAGTGGAGTGGACAAAGGAGGACCAAAACATGAATTCT	1899
Db	1888	GAGTTCAAGATCCAGATACAGTACAGTGGACACGAGGAGGCGGCCCAACTGAATTCT	1947
QY	1900	ACGGCCCAAGAGGTTTACGAAATTTAGTTAGTCCCATTAACACTTCTCCAGAGGCGGAA	1959
Db	1948	ACAACCCGAAGGCGCTATGAAAGTTTGGTTAGCCCAAGTACACCTCTTCTCAGGCTGAC	2007
QY	1960	ATCCTTTTTTGAGATATCCCAAGAAAAACGCTTCCGGAAGTTTACAGACATGATGTT	2019
Db	2008	GTCCCTTTTGAAGATGTCCCAAGAGAAACGCTTCCCAAGATACAGGACACATGATGA	2067
QY	2020	AACTCAACAGGAGGCGCCAGGAAGAGTGAATAATTCCTCTGGTAAATATTTTCACCTCCCT	2079
Db	2068	AATGCAACAGGAGATTCACAGAGGAGTGAATAATCCCGGGTAAATATTTTCACCTCCCT	2127
QY	2080	CCAAAAGACCGCGTGTGAGTCTCAATACCTTGGATTTCGAACATGGAACATGAGACATC	2139
Db	2128	CCAAAAGAGGCGCAGGTGAGGCTGAGCAACTTGGATTTCGAACATGGAACATGAGACATC	2187
QY	2140	ACTTTGAAAGGATACAAATTTGTCCAAGTACGCTTGTGAGATCAATTTCTGATGAACATCA	2199

```

QY 3280 GCATATAAGGACAAACAAATATAGTTTGAATCATGGGAGAGAAATCGCTTTT 3339
Db 3265 GCCTATAAAGACAAACAAATACAGTTTGAATCATGGGAGAGAAATCGCTTTC 3324
QY 3340 AAAATGATTCGTACCAACCTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAAC 3399
Db 3325 AGATGATACGACCAACATGTTCTCATGTGGTGGTGGTGGATGACATCAGAAAAAC 3384
QY 3400 CTTAGGAAGTTGTTGTCCTGAATGACACAAATGACACCAATCAFAAAGATGCTCAGACA 3459
Db 3385 CCCAGGAAGTTCGTTGTCTGAATGACAAATGACACCAATCAFAAAGATGCTCAGACA 3444
QY 3460 GTGAAGCTGTTCTCAGGACACTTATGATCCATGTTCCCATACCTTCCCAATTTGAA 3519
Db 3445 GTGAAGCTGTTCTCAGGACACTTATGATCCATGTTCCCATACCTTCCCAATTTGAG 3504
QY 3520 CTGCCAAGAGATGTCGAAACCGTTTCTTCATATGCATGAGCTGCAGGAATGGAGGCT 3579
Db 3505 CTGCCAAGAGATGTCGAAACCGCTTCTGCACATGCATGAGCTGCAGGAATGGAGGCT 3564
QY 3580 TATCGAGACAAATTTGAAGT 3598
Db 3565 TATCGAGACAAAGCTGAAGT 3583

```

RESULT 9

```

US-09-864-761-10195/c
; Sequence 10195, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomic-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408

```

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 4917
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10195
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005409.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; US-09-864-761-10195

Query Match 30.6%; Score 1103.2; DB 10; Length 1962;
Best Local Similarity 99.7%; Pred. No. 5.9e-308;
Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1581 AGATCATTTTCATGAATTTGTATAAAGTGATCCTTCTCCCAAACACGACTCAGCTATATTAT 1640
Db 1364 AGATCATTTTCATGAATTTGTATAAAGTGATCCTTCTCCCAAACACGACTCAGCTATATTAT 1305
QY 1641 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTGCGAAGATGAGCCAAAAGAGAGTTGA 1700
Db 1304 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTGCGAAGATGAGCCAAAAGAGAGTTGA 1245
QY 1701 AGGTGCTATAGTGACAAATCCAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1760
Db 1244 AGGTGCTATAGTGACAAATCCAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1185
QY 1761 CATCCACTCATATGACAGTGAATGATGCCACCAATACATTTTAACTCTCAGCTT 1820
Db 1184 CATCCACTCATATGACAGTGAATGATGCCACCAATACATTTTAACTCTCAGCTT 1125
QY 1821 TCAAAATACAAACGATGAAGAGTTTCAAAATGCAGATAACAGTGGAGTGACACAAAGGA 1880
Db 1124 TCAAAATACAAACGATGAAGAGTTTCAAAATGCAGATAACAGTGGAGTGACACAAAGGA 1065
QY 1881 GGGACCAAACTGAATTTCTACGGCCCAAGAGGTTAGAAAATTTAGTTAGTCCATAAC 1940
Db 1064 GGGACCAAACTGAATTTCTACGGCCCAAGAGGTTAGAAAATTTAGTTAGTCCATAAC 1005
QY 1941 ACTTCTTCCAGAGCGGAAATCCCTTTTGGAGGATATTCCTCCAAAGAAACGCTTCCCGAA 2000
Db 1004 ACTTCTTCCAGAGCGGAAATCCCTTTTGGAGGATATTCCTCCAAAGAAACGCTTCCCGAA 945
QY 2001 GTTTAAGAGACATGATGTTAACTCAACAAGAGAGAGCCAGAGAGTGAATAATTCCTCT 2060
Db 944 GTTTAAGAGACATGATGTTAACTCAACAAGAGAGAGCCAGAGAGTGAATAATTCCTCT 885
QY 2061 GGTAAATATTTCTACTCTCTCCAAAGAGAGGCTTGAAGTCTCAATACCTTGGATTGCA 2120
Db 884 GGTAAATATTTCTACTCTCTCCAAAGAGAGGCTTGAAGTCTCAATACCTTGGATTGCA 825
QY 2121 ACTGGAACATGAGACATCACTTTGAAAGGATACAAATTTGTCAGGTCAGCTTGTCTGAG 2180
Db 824 ACTGGAACATGAGACATCACTTTGAAAGGATACAAATTTGTCAGGTCAGCTTGTCTGAG 765
QY 2181 ATCATTTCTGATGAACCTCACAGCATGCTAAATAAATAAATAAATAAATAAATAAATAA 2240
Db 764 ATCATTTCTGATGAACCTCACAGCATGCTAAATAAATAAATAAATAAATAAATAAATAA 705
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAACACAGTTTCATATAAAGCATCTTGCCTAA 2300
Db 704 AACAAATGACAGTTTGGTGGCTCCACAGGAAACACAGTTTCATATAAAGCATCTTGCCTAA 645
QY 2301 CAGCTTAGGAGTGTCTGAAACATGTCAGAGGTTTCTTCTTCTGAGTGTGTTAAAGT 2360

```

Db 644 CAGCTTAGGAGTGTCTGAAGATTCAGAGGTTGACTTTCTCGAGTGAGTGAAGT 585  
QY 2361 GAATGTCATGACCGGTCAGATCCACCCCTGGACTTGGAGACACAGCAAGATTTAG 2420  
Db 584 GAATGTCATGACCGGTCAGATCCACCCCTGGACTTGGAGACACAGCAAGATTTAG 525  
QY 2421 AGTGGAACTCACACCCAAACCATAGCGGAAATGTGACAAAGAAAGCCCCCATC 2480  
Db 524 AGTGGAACTCACACCCAAACCATAGCGGAAATGTGACAAAGAAAGCCCCCATC 465  
QY 2481 TCTGATTTCTCAGTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA 2540  
Db 464 TCTGATTTCTCAGTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA 405  
QY 2541 AGAAGACAGTACAGTGGAAAGCCAGTGAAGAAATTCAGAAATCACATAGCGTTACTGAAGTTTACT 2600  
Db 404 AGAAGACAGTACAGTGGAAAGCCAGTGAAGAAATTCAGAAATCACATAGCGTTACTGAAGTTTACT 345  
QY 2601 TGGAAAGAGCTGCAGCATACACAGATAGTTTACCTGGGCTTTTGGCCATGGGAGAAAAA 2660  
Db 344 TGGAAAGAGCTGCAGCATACACAGATAGTTTACCTGGGCTTTTGGCCATGGGAGAAAAA 285  
QY 2661 AAAGTATTTCTAGATCTCTCGAGAA 2688  
Db 284 AAAGTATTTCTAGATCTCTCGAGAA 287

## RESULT 10

US-09-864-761-26835/c  
; Sequence 26835, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 26835  
; LENGTH: 1125  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005409.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.93  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4  
; OTHER INFORMATION: EST\_HUMAN HIT: A1499228.1, EVALUE 0.00e+00  
; OTHER INFORMATION: NT HIT: AB033034.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P36166, EVALUE 3.30e-01  
US-09-864-761-26835

Query Match 30.5%; Score 1101.8; DB 10; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-307;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1581 AGATCATTTTCATCAATTTGTATAAAGTGATCCTTCTCCCAAAACACACTACTATATAT 1640  
Db 1105 AGATCATTTTCATCAATTTGTATAAAGTGATCCTTCTCCCAAAACACACTACTATATAT 1046  
QY 1641 TCCAAAGAGTGAATGCTGCTCTTATTTCAGCTTTTCAGAGTAGTCCCAAAAGAGGAGTTGA 1700  
Db 1045 TCCAAAGAGTGAATGCTGCTCTTATTTCAGCTTTTCAGAGTAGTCCCAAAAGAGGAGTTGA 986  
QY 1701 AGTGGCTTAGTGACAAATCCAAATTAATTCGACATGCTTCTATTGCCAAAGTGGAAAAAC 1760  
Db 985 AGTGGCTTAGTGACAAATCCAAATTAATTCGACATGCTTCTATTGCCAAAGTGGAAAAAC 926  
QY 1761 CATCCACCTCATATAATGCACAGTGGAAATGATGCCACCAACATACATTTTAATCTCACGTT 1820  
Db 925 CATCCACCTCATATAATGCACAGTGGAAATGATGCCACCAACATACATTTTAATCTCACGTT 866  
QY 1821 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAGGGA 1880  
Db 865 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAGGGA 806  
QY 1881 GGGACCAAACTGAATTTCTACGGCCCGAGAGGGTTACGAAAAATTTAGTTAGTCCCATAC 1940  
Db 805 GGGACCAAACTGAATTTCTACGAGCCCGAGAGGGTTACGAAAAATTTAGTTAGTCCCATAC 746  
QY 1941 ACTTCTTCCAGAGCGGGAATCCTTTTTCAGGATATTCCTCAAGAAACACGCTTCCCGAA 2000  
Db 745 ACTTCTTCCAGAGCGGGAATCCTTTTTCAGGATATTCCTCAAGAAACACGCTTCCCGAA 686  
QY 2001 GTTTAGACATGATGTTTAACTCAACAGGAGAGCCAGGAGAGGTTGAAAAATTTCCCT 2060  
Db 685 GTTTAGACATGATGTTTAACTCAACAGGAGAGCCAGGAGAGGTTGAAAAATTTCCCT 626  
QY 2061 GGTAAATATTTTCACTCTTCCAAAGACGCCCGAGTTGAGTCTCAATACCTTGGATTTGCA 2120  
Db 625 GGTAAATATTTTCACTCTTCCAAAGACGCCCGAGTTGAGTCTCAATACCTTGGATTTGCA 566  
QY 2121 ACTGGAACATGCAGACATCACTTTTCAAGAGGATACAAATTTGTCACAGTCCAGCTTCTCAG 2180  
Db 565 ACTGGAACATGCAGACATCACTTTTCAAGAGGATACAAATTTGTCACAGTCCAGCTTCTCAG 506  
QY 2181 ATCATTTCTGTGAACTCACAGCATGCTTAAATAAAAAATCAAGCTATAATAACAGATGA 2240  
Db 505 ATCATTTCTGTGAACTCACAGCATGCTTAAATAAAAAATCAAGCTATAATAACAGATGA 446  
QY 2241 AACAAATGACAGTTTGGTGGCTTCCACAGGAAACAGGTTTCATAAACGATCTTGGCCAAA 2300  
|||||





; ORGANISM: Rattus rattus  
US-09-895-072-12

Query Match 6.5%; Score 233.4; DB 10; Length 460;  
Best Local Similarity 82.7%; Pred. No. 6.3e-57;  
Matches 267; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 666 ATTTCACCAACATTCAGGAAACAATCAACTAAACAAACAAATTCAGGAAATCTTTC 725  
DB 1 ATTCCCAACCAATTCAGGAGAGAGTCACTGAGACAAACAACTGCCAGAAATCTTTC 60  
QY 726 CTCCTAAAGTCAACTGTTCAGTGTGATTTCAGAGGCGAGTGTAGCGCTCTTAAACACTGAA 785  
DB 61 TTCTAAATAAATACTGTTCAGCTGTACTCGGAGGCCGCGCTCTTCTGAAATGAA 120  
QY 786 TAACCCCAAGGATTTTCAAGATTTGAATAGCAAACTAAGAGAAACATGACCATGATGG 845  
DB 121 TAACCCCAAGGATTTTCCCGAGGTGAACAGAGAGCAAGAGAAACATGAGCATGATGG 180  
QY 846 AAAGAACTGACCATAGCTCTCATATTTATTTATGGGATCTGAGCGCCATCAGCCAGTC 905  
DB 181 GAAGGAACTGACCATAGCTCTCATATTTATTTATGGGATCTGAGCGCCATCAGCCAGTC 240  
QY 906 TAAGCAGGATGAAGACATCTCTCCAGTCTGTTTGAAGATAACGAAGAACTGAGGTACTC 965  
DB 241 CAAGCAGGATGAAGATGTGTCTCCAGCGCTTCGAGGAGATAACGAAGAGCTGAGGTACTC 300  
QY 966 ATTGCGATCTATCGAGGAGCATG 988  
DB 301 ACTGAGATCTATCGAGAGACATG 323

## RESULT 14

US-09-986-552-12  
; Sequence 12, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 12  
; LENGTH: 460  
; TYPE: DNA  
; ORGANISM: Rattus rattus  
US-09-986-552-12

Query Match 6.5%; Score 233.4; DB 10; Length 460;  
Best Local Similarity 82.7%; Pred. No. 6.3e-57;  
Matches 267; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 666 ATTTCACCAACATTCAGGAAACAATCAACTAAACAAACAAATTCAGGAAATCTTTC 725  
DB 1 ATTCCCAACCAATTCAGGAGAGAGTCACTGAGACAAACAACTGCCAGAAATCTTTC 60  
QY 726 CTCCTAAAGTCAACTGTTCAGTGTGATTTCAGAGGCGAGTGTAGCGCTCTTAAACACTGAA 785  
DB 61 TTCTAAATAAATACTGTTCAGCTGTACTCGGAGGCCGCGCTCTTCTGAAATGAA 120  
QY 786 TAACCCCAAGGATTTTCAAGATTTGAATAGCAAACTAAGAGAAACATGACCATGATGG 845  
DB 121 TAACCCCAAGGATTTTCCCGAGGTGAACAGAGAGCAAGAGAAACATGAGCATGATGG 180  
QY 846 AAAGAACTGACCATAGCTCTCATATTTATTTATGGGATCTGAGCGCCATCAGCCAGTC 905  
DB 181 GAAGGAACTGACCATAGCTCTCATATTTATTTATGGGATCTGAGCGCCATCAGCCAGTC 240

QY 906 TAAGCAGGATGAAGACATCTCTCCAGTCTGTTTGAAGATAACGAAGAACTGAGGTACTC 965  
DB 241 CAAGCAGGATGAAGATGTGTCTCCAGCGCTTCGAGGATAACGAAGAGCTGAGGTACTC 300  
QY 966 ATTGCGATCTATCGAGAGGAGCATG 988  
DB 301 ACTGAGATCTATCGAGAGACATG 323

## RESULT 15

US-09-864-761-10706/c  
; Sequence 10706, Application us/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 10706  
; LENGTH: 459  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005409.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.89  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.93  
US-09-864-761-10706

Query Match	5.88;	Score 208.8;	DB 10;	Length 459;
Best Local Similarity	99.18;	Pred. No. 8.1e-50;		
Matches 210;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	334	AGTGAGAAGCAGTTAGAGTGTTCCTAACACACTGCAATTAAGGTGCCAATGCTTGCTCG	393
Db	438	AGTGAGAAGCAGTTAGAGTGTTCCTAACACACTGCAATTAAGGTGCCAATGCTTGCTCG	379
Qy	394	GACCCAGCCCTGCCAGCCACATCACCTGAAGGACCTGCCATCTCTTTATCCTTCTTTT	453
Db	378	GACCCAGCCCTGCCAGCCACATCACCTGAAGGACCTGCCATCTCTTTATCCTTCTTTT	319
Qy	454	CATTCTGCCAGTGACATTTTCAATGTTGCCAAACCAAAACCCCTTCTACCAATGCTCA	513
Db	318	CATTCTGCCAGTGACATTTTCAATGTTGCCAAACCAAAACCCCTTCTACCAATGCTCA	259
Qy	514	GTTGTTGTTTTTGACAGTACTAAGGATGTTGA	545
Db	258	GTTGTTGTTTTTGACAGTACTAAGGATGTTAA	227

Search completed: April 13, 2003, 03:03:45  
Job time : 225 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:22:54 ; Search time 33.8869 Seconds  
(without alignments)  
2163.137 Million cell updates/sec

Title: US-10-023-888-2  
Perfect score: 6340  
Sequence: 1 METDTLLVLLVWPGSTG.....NRFLHMLHQLQWRAIDKDK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published\_Applications\_AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pcp.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pcp.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4630	73.0	928	10	US-09-895-072-1
2	4630	73.0	928	10	US-09-896-552-1
3	3639	57.4	908	10	US-09-895-072-15
4	3639	57.4	908	10	US-09-896-552-15
5	1893	29.9	367	10	US-09-864-761-42893
6	1487	23.5	328	10	US-09-895-072-2
7	1487	23.5	328	10	US-09-896-552-2
8	1465	23.1	328	10	US-09-895-072-8
9	1465	23.1	328	10	US-09-896-552-8
10	742	11.7	502	10	US-09-895-072-13
11	742	11.7	502	10	US-09-896-552-13
12	503	7.9	113	10	US-09-895-072-11
13	503	7.9	113	10	US-09-896-552-11
14	487	7.7	112	9	US-10-079-623-349
15	352	5.6	68	10	US-09-864-761-43413
16	150.5	2.4	1639	9	US-10-087-464-10
17	145.5	2.3	2444	10	US-09-944-849-2
18	145	2.3	28	10	US-09-895-072-27
19	145	2.3	28	10	US-09-896-552-27

## ALIGNMENTS

RESULT 1  
US-09-895-072-1  
; Sequence 1, Application US/09895072  
; Patent No. US20020025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-895-072-1

Query Match 73.0%; Score 4630; DB 10; Length 928;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Qy 35 SRDQYHVLFDYSYRNIAGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVRQMEEQ 94  
Db 45 SRDQYHVLFDYSYRNIAGKSFQNRCLCLPMPIDVVYTWVNGTDLLELLKELQOVRQMEEQ 104  
Qy 95 KAMREILGKNTTEPTKSEKOLECLTHCIKVPMLVLDLPALPANITLKDLPSPYPSFSA 154  
Db 105 KAMREILGKNTTEPTKSEKOLECLTHCIKVPMLVLDLPALPANITLKDLPSPYPSFSA 164  
Qy 155 SDIFNVAKPKNPSTNVSVVFDSTKDVDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLM 214  
Db 165 SDIFNVAKPKNPSTNVSVVFDSTKDVDAHSGLLKGNRSQTVWRGYLTDDKEVPGVLVLM 224  
Qy 215 QDLAFSLGFPPTFTETNOLKTKLPENLSKVKLLQLYSEASVALLKNNPKDFOLNKQT 274  
Db 225 QDLAFSLGFPPTFTETNOLKTKLPENLSKVKLLQLYSEASVALLKNNPKDFOLNKQT 284

QY 275 KKNWIDGKELTISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVWR 334  
DB 285 KKNWIDGKELTISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVWR 344  
QY 335 NFIYVINGQIPSWLNDNPRVTIVTHQDVFNRLSHLPTFSSPAESHVHRIEGLSOKFIY 394  
DB 345 NFIYVINGQIPSWLNDNPRVTIVTHQDVFNRLSHLPTFSSPAESHVHRIEGLSOKFIY 404  
QY 395 LNDVDMFGKDWMPDDFYSHSGQKVYLTWPVPCNAEGCGPGSWIKDGYCDKACNNSACDWD 454  
DB 405 LNDVDMFGKDWMPDDFYSHSGQKVYLTWPVPCNAEGCGPGSWIKDGYCDKACNNSACDWD 464  
QY 455 GDCSGNSGGRYIAGGGTGSIGVQPGQGGINSVSYNQCANSWLADKFCDOACN 514  
DB 465 GDCSGNSGGRYIAGGGTGSIGVQPGQGGINSVSYNQCANSWLADKFCDOACN 524  
QY 515 VLSGCFDAGDCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSDNP 574  
DB 525 VLSGCFDAGDCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSDNP 584  
QY 575 IIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTREGPKLNST 634  
DB 585 IIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTREGPKLNST 644  
QY 635 AOKGYENLVSPITLLPEAEILFEDIPKEKRPKFRHDVNSTRAQAEVKIPLVNISSLP 694  
DB 645 AOKGYENLVSPITLLPEAEILFEDIPKEKRPKFRHDVNSTRAQAEVKIPLVNISSLP 704  
QY 695 KDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAHAKIKNOAIITDETNDLSVA 754  
DB 705 KDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAHAKIKNOAIITDETNDLSVA 764  
QY 755 POEKQVHKSILPNSLGVSRERLQRTFFPAVSVKVNHDQGNPPDLLETTARFRVETHQK 814  
DB 765 POEKQVHKSILPNSLGVSRERLQRTFFPAVSVKVNHDQGNPPDLLETTARFRVETHQK 824  
QY 815 TIGGNVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 874  
DB 825 TIGGNVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 884  
QY 875 TDSYLGFLPWEKKYFYDLDDEESLKTQLAYFTDSKNR 915  
DB 885 TDSYLGFLPWEKKYFYDLDDEESLKTQLAYFTDSKNTR 925

## RESULT 2

US-09-986-552-1  
; Sequence 1, Application us/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US7D1V  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-986-552-1

Query Match 73.0%; Score 4630; DB 10; Length 928;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 875; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 35 SRDQYHVLFDSDYRDNIAGKSFQNRILCLPMPIDVVTWVNGTDLLELKLQVREQMEERQ 94  
|||||

DB 45 SRDQYHVLFDSDYRDNIAGKSFQNRILCLPMPIDVVTWVNGTDLLELKLQVREQMEERQ 104  
QY 95 KAMREILGNKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITIKDLPISYFSFHA 154  
DB 105 KAMREILGNKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITIKDVPISYFSFHA 164  
QY 155 SDIFNVAKPNESTNVVVVFDSTKDEDAHSGLLKGNRSQRTVWRGYLTDTKEVPGVLV 214  
DB 165 SDIFNVAKPNESTNVVVVFDSTKDEDAHSGLLKGNRSQRTVWRGYLTDTKEVPGVLV 224  
QY 215 QDLAFLSGPPPTFKETNQIKTLPENLSKVLLQIYSEASVALLKLNPKDFQELNQT 274  
DB 225 QDLAFLSGPPPTFKETNQIKTLPENLSKVLLQIYSEASVALLKLNPKDFQELNQT 284  
QY 275 KKNWTDGRELATISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVWR 334  
DB 285 KKNWTDGRELATISPAYLLWDLISAISQSKQDEDISASRFEDNEELRYSLRISIERHAPVWR 344  
QY 335 NFIYVINGQIPSWLNDNPRVTIVTHQDVFNRLSHLPTFSSPAESHVHRIEGLSOKFIY 394  
DB 345 NFIYVINGQIPSWLNDNPRVTIVTHQDVFNRLSHLPTFSSPAESHVHRIEGLSOKFIY 404  
QY 395 LNDVDMFGKDWMPDDFYSHSGQKVYLTWPVPCNAEGCGPGSWIKDGYCDKACNNSACDWD 454  
DB 405 LNDVDMFGKDWMPDDFYSHSGQKVYLTWPVPCNAEGCGPGSWIKDGYCDKACNNSACDWD 464  
QY 455 GDCSGNSGGRYIAGGGTGSIGVQPGQGGINSVSYNQCANSWLADKFCDOACN 514  
DB 465 GDCSGNSGGRYIAGGGTGSIGVQPGQGGINSVSYNQCANSWLADKFCDOACN 524  
QY 515 VLSGCFDAGDCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSDNP 574  
DB 525 VLSGCFDAGDCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGEVAYSDNP 584  
QY 575 IIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTREGPKLNST 634  
DB 585 IIRHASIAKWKTIHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTREGPKLNST 644  
QY 635 AOKGYENLVSPITLLPEAEILFEDIPKEKRPKFRHDVNSTRAQAEVKIPLVNISSLP 694  
DB 645 AOKGYENLVSPITLLPEAEILFEDIPKEKRPKFRHDVNSTRAQAEVKIPLVNISSLP 704  
QY 695 KDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAHAKIKNOAIITDETNDLSVA 754  
DB 705 KDAQSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAHAKIKNOAIITDETNDLSVA 764  
QY 755 POEKQVHKSILPNSLGVSRERLQRTFFPAVSVKVNHDQGNPPDLLETTARFRVETHQK 814  
DB 765 POEKQVHKSILPNSLGVSRERLQRTFFPAVSVKVNHDQGNPPDLLETTARFRVETHQK 824  
QY 815 TIGGNVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 874  
DB 825 TIGGNVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAEHIGVTEVLLGRKLOHY 884  
QY 875 TDSYLGFLPWEKKYFYDLDDEESLKTQLAYFTDSKNR 915  
DB 885 TDSYLGFLPWEKKYFYDLDDEESLKTQLAYFTDSKNTR 925

## RESULT 3

US-09-895-072-15  
; Sequence 15, Application us/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCONT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 15  
 ; LENGTH: 908  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-895-072-15

Query Match 57.4%; Score 3639; DB 10; Length 908;  
 Best Local Similarity 78.7%; Pred. No. 5.1e-258;  
 Matches 694; Conservative 60; Mismatches 106; Indels 22; Gaps 7;

Qy	35	SRDQYHVLFDSDYRNIAGKSFQNRCLCPMPIDVVTWNGTDLLELLKELQOVRHEMEEQ	94
Db	45	SRDQYHVLFDSDYRNIAGKSFQNRCLCPMPIDVVTWNGTDLLELLKELQOVRHEMEEQ	104
Qy	95	KAMREILGKNTTEPTKKSEKOLECLLTHCIKVPMLVLDPALPANITLKDPLSPYSPFHAS	154
Db	105	KAMREILGKNTTEPTKKSEKOLECLLTHCIKVPMLVLDPALPANITLKDPLSPYSPFHAS	164
Qy	155	SDIFNVAKPNPSTNVSVVDFSTKDVDAHSGLLKGNRSQTVWVRYLTDDKEVPGGLVLM	214
Db	165	SDIFNVAKPNPSTNVSVVDFSTKDVDAHSGLLKGNRSQTVWVRYLTDDKEVPGGLVLM	224
Qy	215	QDLAFSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKO	273
Db	225	QDLAFSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKO	284
Qy	274	TKKNTIDGKELTISPAYLLWDLAISQSKODEDASASREFEDNEELRYSLRISERHAPW	333
Db	285	TKKNTIDGKELTISPAYLLWDLAISQSKODEDASASREFEDNEELRYSLRISERHAPW	344
Qy	334	RNIFVINGQIPSWNLNDNPRVTIVTHQDFVFNLSHLPTSSPAIESHVRIEGLSQKFI	393
Db	345	RNIFVINGQIPSWNLNDNPRVTIVTHQDFVFNLSHLPTSSPAIESHVRIEGLSQKFI	404
Qy	394	YLNDDVDFGKDWDPDDFYSHSGKGVLTWVPVPCAGCGPWSWIKDGYCDKACNNSACDW	453
Db	405	YLNDDVDFGKDWDPDDFYSHSGKGVLTWVPVPCAGCGPWSWIKDGYCDKACNNSACDW	464
Qy	454	DGDCSGNSGGSRYIAGGGTGTSGVGPWFQGGGINSYCNQGCANSWLADKFCDOAC	513
Db	465	DGDCSGNSGGSRYIAGGGTGTSGVGPWFQGGGINSYCNQGCANSWLADKFCDOAC	524
Qy	514	NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYVVPVPCAGCGPWSWIKDGYCDKACNNSACDW	573
Db	525	NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYVVPVPCAGCGPWSWIKDGYCDKACNNSACDW	584
Qy	574	PIIRHASIANKWKTILHIMSGMNATTIHFNLTFONTNDEEFKMOITVEVDTRREGPKLNS	633
Db	585	PIIRHASIANKWKTILHIMSGMNATTIHFNLTFONTNDEEFKMOITVEVDTRREGPKLNS	644
Qy	634	TAQKAYENLVSPITLLPEAEILFEDIPKRRFPKRRHVDNSTRRAQEEVKIPRVNISLL	693
Db	645	TAQKAYENLVSPITLLPEAEILFEDIPKRRFPKRRHVDNSTRRAQEEVKIPRVNISLL	704
Qy	694	PKDAQSLNTLDLQLEHGDITLKYNLKSKALLRSLFNLMSQAHAKIKQAIITDETNDLSV	753
Db	705	PKDAQSLNTLDLQLEHGDITLKYNLKSKALLRSLFNLMSQAHAKIKQAIITDETNDLSV	763
Qy	754	APQKQVHKSILPNSLGVSRLOLTPPAVSVKVNQGHGQGNPLDLETTARFVETHQ	813
Db	764	APQKQVHKSILPNSLGVSRLOLTPPAVSVKVNQGHGQGNPLDLETTARFVETHQ	816
Qy	814	KTIGNVTKPPSLIPLVPSQMTKEKTKGKENSMEENAEHNGVTEVLLGRKLQH	873
Db	817	KTIGNVTKPPSLIPLVPSQMTKEKTKGKENSMEENAEHNGVTEVLLGRKLQH	885
Qy	874	YTDVSLGFLPWEKKYFLLDDEESLKTQLAYETDSEKNNAR	915
Db	866	YTDVSLGFLPWEKKYFLLDDEESLKTQLAYETDSEKNNAR	905

RESULT 4

US-09-986-552-15  
 ; Sequence 15, Application US/09986552  
 ; Patent No. US20020150981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CANFIELD, William  
 ; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
 ; FILE REFERENCE: 215089US77DIV  
 ; CURRENT APPLICATION NUMBER: US/09/986,552  
 ; CURRENT FILING DATE: 2001-11-09  
 ; PRIOR APPLICATION NUMBER: 09/635,872  
 ; PRIOR FILING DATE: 2000-08-10  
 ; PRIOR APPLICATION NUMBER: 60/153,831  
 ; PRIOR FILING DATE: 1999-09-14  
 ; NUMBER OF SEQ ID NOS: 52  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 15  
 ; LENGTH: 908  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-986-552-15

Query Match 57.4%; Score 3639; DB 10; Length 908;  
 Best Local Similarity 78.7%; Pred. No. 5.1e-258;  
 Matches 694; Conservative 60; Mismatches 106; Indels 22; Gaps 7;

Qy	35	SRDQYHVLFDSDYRNIAGKSFQNRCLCPMPIDVVTWNGTDLLELLKELQOVRHEMEEQ	94
Db	45	SRDQYHVLFDSDYRNIAGKSFQNRCLCPMPIDVVTWNGTDLLELLKELQOVRHEMEEQ	104
Qy	95	KAMREILGKNTTEPTKKSEKOLECLLTHCIKVPMLVLDPALPANITLKDPLSPYSPFHAS	154
Db	105	KAMREILGKNTTEPTKKSEKOLECLLTHCIKVPMLVLDPALPANITLKDPLSPYSPFHAS	164
Qy	155	SDIFNVAKPNPSTNVSVVDFSTKDVDAHSGLLKGNRSQTVWVRYLTDDKEVPGGLVLM	214
Db	165	SDIFNVAKPNPSTNVSVVDFSTKDVDAHSGLLKGNRSQTVWVRYLTDDKEVPGGLVLM	224
Qy	215	QDLAFSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKO	273
Db	225	QDLAFSGFPPTFKETNQLTKLPEN-LSSKVKLLQLYSEASVALLKNNPKDFQELNKO	284
Qy	274	TKKNTIDGKELTISPAYLLWDLAISQSKODEDASASREFEDNEELRYSLRISERHAPW	333
Db	285	TKKNTIDGKELTISPAYLLWDLAISQSKODEDASASREFEDNEELRYSLRISERHAPW	344
Qy	334	RNIFVINGQIPSWNLNDNPRVTIVTHQDFVFNLSHLPTSSPAIESHVRIEGLSQKFI	393
Db	345	RNIFVINGQIPSWNLNDNPRVTIVTHQDFVFNLSHLPTSSPAIESHVRIEGLSQKFI	404
Qy	394	YLNDDVDFGKDWDPDDFYSHSGKGVLTWVPVPCAGCGPWSWIKDGYCDKACNNSACDW	453
Db	405	YLNDDVDFGKDWDPDDFYSHSGKGVLTWVPVPCAGCGPWSWIKDGYCDKACNNSACDW	464
Qy	454	DGDCSGNSGGSRYIAGGGTGTSGVGPWFQGGGINSYCNQGCANSWLADKFCDOAC	513
Db	465	DGDCSGNSGGSRYIAGGGTGTSGVGPWFQGGGINSYCNQGCANSWLADKFCDOAC	524
Qy	514	NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYVVPVPCAGCGPWSWIKDGYCDKACNNSACDW	573
Db	525	NVLSCGFDAGCGQDHFHELYKVTLLPNQTHYVVPVPCAGCGPWSWIKDGYCDKACNNSACDW	584
Qy	574	PIIRHASIANKWKTILHIMSGMNATTIHFNLTFONTNDEEFKMOITVEVDTRREGPKLNS	633
Db	585	PIIRHASIANKWKTILHIMSGMNATTIHFNLTFONTNDEEFKMOITVEVDTRREGPKLNS	644
Qy	634	TAQKAYENLVSPITLLPEAEILFEDIPKRRFPKRRHVDNSTRRAQEEVKIPRVNISLL	693
Db	645	TAQKAYENLVSPITLLPEAEILFEDIPKRRFPKRRHVDNSTRRAQEEVKIPRVNISLL	704
Qy	694	PKDAQSLNTLDLQLEHGDITLKYNLKSKALLRSLFNLMSQAHAKIKQAIITDETNDLSV	753
Db	705	PKDAQSLNTLDLQLEHGDITLKYNLKSKALLRSLFNLMSQAHAKIKQAIITDETNDLSV	763

Qy	754	APQEKQVHKSIPLNSLGVSERLQRLTFPAVSVKVNCHDQONPPDLETTARFVETHQ	813
Db	764	VPQENPSHR--PHGRAGEHRSERWAPAEVTVKGRDHALNPPPVLETNARU-----AQ	816
Qy	814	KTGGNNVTKKPSLPIVPLESQMTKEKKTGKENSMEENAEHIGVTEVLLGRKLOH	873
Db	817	PTLGVTVSKENSLPIVPPESHLP-----KEESDRAEGNA---VPVKELVPGRLQ-	865
Qy	874	YTSYLGFLPWEEKYFLLDDEESLKTQLAYFTOSKNRAR	915
Db	866	--ONYGFGLPWEEKYFQDLDDEESLKTQLAYFTDKHTGR	905

```

RESULT 5
US-09-864-761-42893
; Sequence 42893, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42893
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005409.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5

```

Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 1098  
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 180  
QY 1099 NKYRFEIMGEEIEAFKMIRTNVSHVVGQLDDIRKNPRKFVCLNDNIDHNHKAQTVKAVL 1158  
Db 181 NKYRFEIMGEEIEAFKMIRTNVSHVVGQLDDIRKNPRKFVCLNDNIDHNHKAQTVKAVL 240  
QY 1159 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 1199  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

RESULT 7  
US-09-986-552-2  
; Sequence 2, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 2  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-986-552-2

Query Match 23.5%; Score 1487; DB 10; Length 328;  
Best Local Similarity 100.0%; Pred. No. 5.1e-101;  
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 919 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 978  
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 60  
QY 979 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 1038  
Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 1098  
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 180  
QY 1099 NKYRFEIMGEEIEAFKMIRTNVSHVVGQLDDIRKNPRKFVCLNDNIDHNHKAQTVKAVL 1158  
Db 181 NKYRFEIMGEEIEAFKMIRTNVSHVVGQLDDIRKNPRKFVCLNDNIDHNHKAQTVKAVL 240  
QY 1159 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 1199  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

RESULT 8  
US-09-895-072-8  
; Sequence 8, Application US/09895072  
; Patent No. US20020025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCNT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-895-072-8

Query Match 23.1%; Score 1465; DB 10; Length 328;  
Best Local Similarity 98.2%; Pred. No. 2.1e-99;  
Matches 276; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 919 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 978  
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 60  
QY 979 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 1038  
Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 1098  
Db 121 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 180  
QY 1099 NKYRFEIMGEEIEAFKMIRTNVSHVVGQLDDIRKNPRKFVCLNDNIDHNHKAQTVKAVL 1158  
Db 181 NKYRFEIMGEEIEAFKMIRTNVSHVVGQLDDIRKNPRKFVCLNDNIDHNHKAQTVKAVL 240  
QY 1159 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 1199  
Db 241 RQFYESMFPPIPSQFELPREYRNRLHMHLEQWRAIRYDKLK 281

RESULT 9  
US-09-986-552-8  
; Sequence 8, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-986-552-8

Query Match 23.1%; Score 1465; DB 10; Length 328;  
Best Local Similarity 98.2%; Pred. No. 2.1e-99;  
Matches 276; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 919 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 978  
Db 1 DTFADSLRYVKNILNSKFGFTSRKVPAPAHMPHMDIRIVMQELQDMFPEEFDKTSFHKVRRHS 60  
QY 979 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 1038  
Db 61 EDMQAFSYFYLLMSAVQPLNISQVFEVDVTDQSGVLSDRIRTLATRIHELPLSLQDLT 120  
QY 1039 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 1098  
Db 1039 GLEHMLINCSKMLPADITQLNNIPPTQESYDNPVTKSLVTNCKPVTDKIHKAYKOK 1098



Db 121 GLEHMLNCSKMLPANITQLNPIPTQAYYDNPVPTKSLVNTCKVPVTDKTHKAYKDK 180  
QY 1099 NKYRPEIMGEEETAFKMTRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTVKAVL 1158  
Db 181 NKYRPEIMGEEETAFKMTRTNVSHVVGQDDIRKNPKRFVCLNDNDHNDKDAQTVKAVL 240  
QY 1159 RDFSMEFPPIPSQFELPREYRNRLHMHLEQEWAYRDKLK 1199  
Db 241 RDFSMEFPPIPSQFELPREYRNRLHMHLEQEWAYRDKLK 281

## RESULT 10

US-09-895-072-13  
; Sequence 13, Application US/09895072  
; Patent No. US2002025550A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM M  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 210119USOCNT  
; CURRENT APPLICATION NUMBER: US/09/895,072  
; CURRENT FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; PRIOR APPLICATION NUMBER: US 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-895-072-13

Query Match 11.7%; Score 742; DB 10; Length 502;  
Best Local Similarity 23.2%; Pred. No. 3.8e-46;  
Matches 197; Conservative 98; Mismatches 181; Indels 372; Gaps 14;

QY 312 RFEDNEELRYSLRSTERHAPWVRNIFVITNGQIPSWMLNDNPRVTIVTHQDVFNLSHLP 371  
Db 4 RFDDKNELRYSLRSLKHAARHVRVITNGQIPSWLDLSYERVTVVPHVPLADPDQLP 63  
QY 372 TFSSPAIESHVHRIEGLSOKFIYLNDDVMFGKDVDPDDFYSHSKGOKVYLTWPVPCNCAEG 431  
Db 64 TFSSAIETFLHRIPKLSRFLYLNDDIFLGAPLYDELITAEAGVRVYQAWVPFGCALD 123  
QY 432 CPGSWIKDGYCDKACNNSACDWDGDCSGNSGGRYIAGGGGTGSGVGPWFQFGGINS 491  
Db 124 CPWTYIGDGCADRHCHNIDACQFDGDCSETGPAS 157  
QY 492 VSYCNOGCANSWLADKFCDAQCNVLSGCFDAGDCQDHFHLYKVILLPNQTHYIIPKGE 551  
Db 158 -----DARVIPPFSKE 167  
QY 552 CLPYFSFAEVAKRGEVAYSDNPIIRHASIANKWKTIIHLMHSGMNATTIHFNLTFQNTN 611  
Db 168 VLE----- 170  
QY 612 DEEFKMQITVEVDTRGPKLNSTAKGYENLVSPITLPEAEILFEDIPKRRFPK 667  
Db 171 -----VQP-AAVPQSRV-----HREFPQGLQ 190  
QY 668 --FKRHDVNSTRAQEEVKIPLVNLISILPKDAQLSNLTDLQLEHGDITLKGYNLSKAL 725  
Db 191 KLFRRSSANF-----KD-----VMRHRNVSTLKE 214  
QY 726 LRSFLMNSOHAKIKNOAITTDETNDSLVAPQEKOVHKSILNLSLGSVSRQLRQTFPAVS 785  
Db 215 LRIVERFNKALMS--LNPELETSSSEPQTQRH----- 247  
QY 786 KVNHDQGNPPDLLETARFRVETHTKTIGGNVTKEKPPSLIVPLESQMTKEKKTIGK 845  
Db 248 -----GL 249

QY 846 EKENSRENEAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKKYFLDILLDEESLKTQIA 905  
Db 250 RKE----- 252  
QY 906 YFTDSKNRARYKRDFTADSLRYVKNILSKFGTSKRYPAHMPHMLDIRVMOELQDMPE 965  
Db 253 ---DFKS---STDIYSHSLIATNMLNRAYGFKARHVAHVGLDKDIDVEMQRRFHQ 305  
QY 966 EFDKTSFHKVRHSEDMQFAFSFYILMSAVQPLNISQVDFEVDTDQSGVLSDSREIRTLAT 1025  
Db 306 QILDTAHOFRAPTDLQYAFAYFSLMSETKVMSVEEIFDEFTDGSATWSREVRTELT 365  
QY 1026 RHEPLSLQDLTGLEHMLNCSK----MLPADITQLNINPTQESYDNPVPTKSLV 1081  
Db 366 RIYQPLDWSAMRYEEVVQNCNRNLGMHLKYDVTVEHSTL--VYREYEDSNLPTITRDLV 423  
QY 1082 TNCKPVTDKIHKAYDKNRYEIMGE--EETAFKMTRTNVSHVVGQDDIRKNPKRFVC 1139  
Db 424 VRCPLLAELAAANFAVRPKYFNHVSFKTSHSNFMMLTSLNTEVYESLDRLRNRKENC 483  
QY 1140 LNDNIDHN 1147  
Db 484 INDNLDAN 491

## RESULT 11

US-09-986-552-13  
; Sequence 13, Application US/09986552  
; Patent No. US20020150981A1  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES  
; FILE REFERENCE: 215089US77DIV  
; CURRENT APPLICATION NUMBER: US/09/986,552  
; CURRENT FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: 09/635,872  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-986-552-13

Query Match 11.7%; Score 742; DB 10; Length 502;  
Best Local Similarity 23.2%; Pred. No. 3.8e-46;  
Matches 197; Conservative 98; Mismatches 181; Indels 372; Gaps 14;

QY 312 RFEDNEELRYSLRSTERHAPWVRNIFVITNGQIPSWMLNDNPRVTIVTHQDVFNLSHLP 371  
Db 4 RFDDKNELRYSLRSLKHAARHVRVITNGQIPSWLDLSYERVTVVPHVPLADPDQLP 63  
QY 372 TFSSPAIESHVHRIEGLSOKFIYLNDDVMFGKDVDPDDFYSHSKGOKVYLTWPVPCNCAEG 431  
Db 64 TFSSAIETFLHRIPKLSRFLYLNDDIFLGAPLYDELITAEAGVRVYQAWVPFGCALD 123  
QY 432 CPGSWIKDGYCDKACNNSACDWDGDCSGNSGGRYIAGGGGTGSGVGPWFQFGGINS 491  
Db 124 CPWTYIGDGCADRHCHNIDACQFDGDCSETGPAS 157  
QY 492 VSYCNOGCANSWLADKFCDAQCNVLSGCFDAGDCQDHFHLYKVILLPNQTHYIIPKGE 551  
Db 158 -----DARVIPPFSKE 167  
QY 552 CLPYFSFAEVAKRGEVAYSDNPIIRHASIANKWKTIIHLMHSGMNATTIHFNLTFQNTN 611  
Db 168 VLE----- 170  
QY 612 DEEFKMQITVEVDTRGPKLNSTAKGYENLVSPITLPEAEILFEDIPKRRFPK 667

Db 171 -----VQP-AAVQSRV-----HRFPQMGLQ 190  
QY 668 --FKRHDVNSTRAQEEVKIPLVNISSLPKDAQLSLNTLDLQLEHCDITLKGYNLSKSL 725  
Db 191 KLFRRSANP-----KD-----VMHRNVSTLKE 214  
QY 726 LRFSLMSQAKIKNOAIITDETNDSLVAPQEQVHKSLPNSLGVSRQLRLTFPAVS 785  
Db 215 LRRIVERENAKLMS---LNPELETSSSEPQTQRH-----247  
QY 786 KVNQHDQOGNPDLLETTARFRVETHQTIGGNVTEKPPSLIPLVESQMTKEKKITGK 845  
Db 248 -----GL 249  
QY 846 EKENSMEENAHENHIGTEVLLGRKLQHYTDSYGLFLPWEKKYFLDLLDEESLKTQLA 905  
Db 250 RKE-----252  
QY 906 YFTDSKNRARKDFTFADSLRYVNIKLSKFGFTSRKVPAAHPMHIDRIVMOELQMPPE 965  
Db 253 ---DFKS---STDYIYSHSLIATNMLNRAYGFKARHVLAVHGVFLIDKDIVEMORRHQ 305  
QY 966 EFDKTSFHKVRUSEDQMFAPSFYFYLMSAVQPLNLSQVDEVDTSQVLSDEIRTIAT 1025  
Db 306 QILDTAQRFRAPTDIQAFAFYSELMSETKVSVEEIPDEFDTDCSATWSDEVRITLT 365  
QY 1026 RIHELPLSLQDLTGLEHMLNLSK---MLPADITOLNIPPTQESYYPDNPVPYTKSLV 1081  
Db 366 RIYQPLDSAMRYFEVVQNCNTRNLGMHLKVDTVEHSTL--VYBRYEDSNLPTITRDLV 423  
QY 1082 TNCKPVTDKIKKAYDKNRYFEIMGE--EEIAFKMIRTNVSHVYGQDLDDIRKNPKFVC 1139  
Db 424 VRCPLAALAAANFAVRPKYHVSFKRTHSHNFMMILSNLTVESLDRLRNRPKFNC 483  
QY 1140 LNDNIDHN 1147  
Db 484 INDNLDAN 491

## RESULT 12

US-09-895-072-11

; Sequence 11, Application US/09895072

; Patent No. US2002025550A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, WILLIAM M

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES

; FILE REFERENCE: 21019USOCNT

; CURRENT APPLICATION NUMBER: US/09/895,072

; CURRENT FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; PRIOR APPLICATION NUMBER: US 09/635,872

; PRIOR FILING DATE: 2000-08-10

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Rattus rattus

US-09-895-072-11

Query Match 7.9%; Score 503; DB 10; Length 113;  
Best Local Similarity 92.5%; Pred. No. 1.2e-29;  
Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 223 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKTOKKNTIDG 282

Db 1 FPPTFKETSQLTKLPENLSSKIKLQLYSEASVALLKLNPKGFPPELNKQTKKNMSISG 60

QY 283 KELTISPAYLLWDLISAQSKODEDISASRFEDNEELRYSLSIERH 329

Db 61 KELAISPAYLLWDLISAQSKODEDVASRFEDNEELRYSLSIERH 107

## RESULT 13

US-09-986-552-11

; Sequence 11, Application US/09986552

; Patent No. US20020150981A1

; GENERAL INFORMATION:

; APPLICANT: CANFIELD, William

; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYOSOMAL HYDROLASES

; FILE REFERENCE: 215089US77DIV

; CURRENT APPLICATION NUMBER: US/09/986,552

; CURRENT FILING DATE: 2001-11-09

; PRIOR APPLICATION NUMBER: 09/635,872

; PRIOR FILING DATE: 2000-08-10

; PRIOR APPLICATION NUMBER: 60/153,831

; PRIOR FILING DATE: 1999-09-14

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11

; LENGTH: 113

; TYPE: PRT

; ORGANISM: Rattus rattus

US-09-986-552-11

Query Match 7.9%; Score 503; DB 10; Length 113;

Best Local Similarity 92.5%; Pred. No. 1.2e-29;

Matches 99; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 223 FPPTFKETNQLTKLPENLSSKVKLLQLYSEASVALLKLNPKDFQELNKTOKKNTIDG 282

Db 1 FPPTFKETSQLTKLPENLSSKIKLQLYSEASVALLKLNPKGFPPELNKQTKKNMSISG 60

QY 283 KELTISPAYLLWDLISAQSKODEDISASRFEDNEELRYSLSIERH 329

Db 61 KELAISPAYLLWDLISAQSKODEDVASRFEDNEELRYSLSIERH 107

## RESULT 14

US-10-079-623-349

; Sequence 349, Application US/10079623

; Patent No. US20020169302A1

; GENERAL INFORMATION:

; APPLICANT: Havukkala, Ilkka J.

; APPLICANT: Glenn, Matthew

; APPLICANT: Grigor, Murray R.

; APPLICANT: Molenaar, Adrian J.

; TITLE OF INVENTION: Compositions isolated from bovine

; FILE REFERENCE: 11000,1044c3

; CURRENT APPLICATION NUMBER: US/10/079,623

; CURRENT FILING DATE: 2002-02-19

; NUMBER OF SEQ ID NOS: 370

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 349

; LENGTH: 112

; TYPE: PRT

; ORGANISM: Bovine

US-10-079-623-349

Query Match 7.7%; Score 487; DB 9; Length 112;  
Best Local Similarity 83.9%; Pred. No. 1.8e-28;  
Matches 94; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 851 RMEENAHNHGVTVELLGRKLQHYTDSYGLFLPWEKKYFLDLLDEESLKTQLAYFTDS 910

Db 1 KIEENANSYPCGNEVPGRKLQYQYDYSYGLFLPWEKKYFODLLDEESLKTQLAYFTDS 60

QY 911 KNARYRDRFADSLRYVNIKLSKFGFTSRKVPAAHPMHIDRIVMOELQDM 962

Db 61 KHTGRQLKDTFADSLRYVNIKLSKFGFTSRKVPAAHPMHIDRIVMOELQDM 112

## RESULT 15

US-09-864-761-43413

; Sequence 43413, Application us/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 43413  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005409.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.89  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.93  
; OTHER INFORMATION: SWISSPROT HIT: P28976, EVALUATE 3.00e-01  
US-09-864-761-43413

Query Match 5.6%; Score 352; DB 10; Length 68;  
Best Local Similarity 100.0%; Pred. No. 6.5e-19;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 EKQLECLLTHCIKYPMLVLPALPANITLKDLPSPFHSASDIFNVAKPNPSTNVSV 172  
Db 1 EKQLECLLTHCIKYPMLVLPALPANITLKDLPSPFHSASDIFNVAKPNPSTNVSV 60

QY 173 VVFDSTKD 180  
Db 61 VVFDSTKD 68  
Search completed: April 13, 2003, 03:31:00  
Job time : 37.8869 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:00:01 ; Search time 0.212801 seconds  
(without alignments)  
3873.042 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rhodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	41	16 Q8VKR0	Q8VKR0 mycobacteri
2	20	100.0	50	10 Q9LUZ2	Q9LUZ2 arabidopsis
3	20	100.0	55	9 Q9ZX12	Q9ZX12 mycobacteri
4	20	100.0	56	5 Q26922	Q26922 trypanosoma
5	20	100.0	58	4 Q9UPG5	Q9UPG5 homo sapien
6	20	100.0	60	2 Q93U42	Q93U42 carsonella
7	20	100.0	60	2 Q93U40	Q93U40 carsonella
8	20	100.0	61	12 Q91GA0	Q91GA0 columbid ci
9	20	100.0	63	2 Q93JW0	Q93JW0 pseudomonas
10	20	100.0	66	2 Q9AP90	Q9AP90 uncultured
11	20	100.0	66	16 Q51760	Q51760 borrelia bu
12	20	100.0	67	16 Q8XGJ9	Q8XGJ9 brucella me
13	20	100.0	68	9 Q38263	Q38263 lactococcus
14	20	100.0	70	16 Q8XRG1	Q8XRG1 ralstonia s
15	20	100.0	71	16 Q9PDH9	Q9PDH9 xyella fas
16	20	100.0	73	10 P93514	P93514 gossypium b

17	20	100.0	74	2 Q47817	Q47817 enterococcu
18	20	100.0	75	16 Q8YSR5	Q8YSR5 anabaena sp
19	20	100.0	75	16 Q8YM57	Q8YM57 anabaena sp
20	20	100.0	75	16 Q8XNX3	Q8XNX3 clostridium
21	20	100.0	77	3 Q9P5V0	Q9P5V0 neurospora
22	20	100.0	78	2 P82548	P82548 streptococc
23	20	100.0	78	9 Q37833	Q37833 bacterioph
24	20	100.0	79	11 P97446	P97446 mus musculu
25	20	100.0	80	16 Q8YXU1	Q8YXU1 anabaena sp
26	20	100.0	81	12 Q914G2	Q914G2 sulfolobus
27	20	100.0	82	10 Q80460	Q80460 arabidopsis
28	20	100.0	85	17 Q8Q3Y8	Q8Q3Y8 pyrococcus
29	20	100.0	86	10 Q9MOC7	Q9MOC7 arabidopsis
30	20	100.0	86	12 Q67406	Q67406 influenzavi
31	20	100.0	86	12 Q67407	Q67407 influenzavi
32	20	100.0	86	12 Q67408	Q67408 influenzavi
33	20	100.0	87	6 Q77640	Q77640 macaca mula
34	20	100.0	87	12 Q91U03	Q91U03 influenza a
35	20	100.0	88	6 Q9BDH8	Q9BDH8 bos taurus
36	20	100.0	88	10 Q9ZV45	Q9ZV45 arabidopsis
37	20	100.0	90	10 Q9LFW8	Q9LFW8 arabidopsis
38	20	100.0	91	12 Q91U02	Q91U02 influenza a
39	20	100.0	92	10 Q9ASJ2	Q9ASJ2 oryza sativ
40	20	100.0	92	17 Q97X43	Q97X43 sulfolobus
41	20	100.0	94	10 Q9FTL0	Q9FTL0 oryza sativ
42	20	100.0	94	12 Q8V6R1	Q8V6R1 halovirus h
43	20	100.0	94	16 Q914Y9	Q914Y9 pseudomonas
44	20	100.0	99	15 Q40235	Q40235 human immun
45	20	100.0	99	15 Q40238	Q40238 human immun

## ALIGNMENTS

## RESULT 1

Q8VKR0	ID	Q8VKR0	PRELIMINARY;	PRT;	41 AA.
AC	Q8VKR0;				
DT	01-MAR-2002 (Tremblrel. 20, Created)				
DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)				
DE	Hypothetical protein MT0159.				
GN	MT0159				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / OSHKOSH;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RA	Bishai W.				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AE006927; AAK44383.1; -				
DR	TIGR; MT0159; -				
KW	Hypothetical protein.				
SQ	SEQUENCE 41 AA; 4434 MW; 4E1DEA437CAA07B9 CRC64;				

Query Match 100.0%; Score 20; DB 16; Length 41;  
Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 26 IEGR 29

## RESULT 2

Q9LUZ2  
ID Q9LUZ2 PRELIMINARY; PRT; 50 AA.  
AC Q9LUZ2  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE Genomic DNA, chromosome 5, p1 clone:M2N1.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20181125; PubMed=10718197;  
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
RT features of the regions of 3,076,755 bp covered by sixty p1 and TAC  
RT clones.";  
RL DNA Res. 7:31-63(2000).  
DR EMBL; AB020755; BAA97334.1;  
SQ SEQUENCE 50 AA; 5958 MW; EDEFF7A687643966 CRC64;  
  
Query Match 100.0%; Score 20; DB 10; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 IEGR 4  
Db 39 IEGR 42  
  
RESULT 3  
Q9ZX12  
ID Q9ZX12 PRELIMINARY; PRT; 55 AA.  
AC Q9ZX12;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)  
DE Gp68.  
OS Mycobacteriophage TM4.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.  
OX NCBI\_TaxID=88870;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20110038; PubMed=10645443;  
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;  
RT "Mycobacteriophage TM4: Genome structure and gene expression.";  
RL Tuber. Lung Dis. 79:63-73(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ford M.E., Stenstrom C., Hendrix R.W., Hatfull G.F.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF068845; AAD17633.1;  
SQ SEQUENCE 55 AA; 5782 MW; F2746364DB90A01F CRC64;  
  
Query Match 100.0%; Score 20; DB 9; Length 55;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 IEGR 4  
Db 51 IEGR 54  
  
RESULT 4  
Q26922  
ID Q26922 PRELIMINARY; PRT; 56 AA.  
AC Q26922;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE G protein (Fragment).  
OS Trypanosoma cruzi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5693;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y;  
RA Lopes U.G.;  
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; M92652; AAA91960.1;  
DR InterPro; IPR001806; Ras\_trnsmfmg.  
DR InterPro; IPR003575; Small\_GTPase.  
DR Pfam; PF00071; ras; 1.  
DR SMART; SM00010; small\_GTPase; 1.  
FT NON\_TER 1 56  
FT NON\_TER 56  
SQ SEQUENCE 56 AA; 6283 MW; B3DBFC0450453726 CRC64;  
  
Query Match 100.0%; Score 20; DB 5; Length 56;  
Best Local Similarity 100.0%; Pred. No. 5.4e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 IEGR 4  
Db 39 IEGR 42  
  
RESULT 5  
Q9UPG5  
ID Q9UPG5 PRELIMINARY; PRT; 58 AA.  
AC Q9UPG5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Serine-threonine kinase (Fragment).  
GN BTAK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BREAST;  
RA Sen S., Zhou H.;  
RT "Partial Genomic Sequence of BTAK.";  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF011467; AAC23448.1;  
DR InterPro; IPR000719; Euk\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Kinase.  
FT NON\_TER 1 58  
FT NON\_TER 58  
SQ SEQUENCE 58 AA; 6812 MW; 30D49CBF5F6B5F86 CRC64;  
  
Query Match 100.0%; Score 20; DB 4; Length 58;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 IEGR 4  
Db 16 IEGR 19  
  
RESULT 6  
Q93U42  
ID Q93U42 PRELIMINARY; PRT; 60 AA.  
AC Q93U42;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE RNA polymerase beta subunit (Fragment).  
 GN RPOB.  
 OS Carsonella ruddii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
 OX NCBI\_TaxID=114186;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;  
 RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts  
 (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons  
 with 16S-23S rDNA-derived phylogeny.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC EMBL; AF268062; AAK55952.1; -.  
 DR InterPro: IPR001572; RNA\_pol\_B.  
 DR Pfam: PF00562; RNA\_pol\_B; 1.  
 DR DNA-directed RNA polymerase; Transcription; Transferase.  
 KW NON\_TER 1  
 FT SEQUENCE 60 AA; 6786 MW; 90BB0074478AFBFD CRC64;  
 SQ

Query Match 100.0%; Score 20; DB 2; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 ||||  
 DB 18 IEGR 21

RESULT 7  
 Q93U40  
 ID Q93U40 PRELIMINARY; PRT; 60 AA.  
 AC Q93U40;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE RNA polymerase beta subunit (Fragment).  
 GN RPOB.  
 OS Carsonella ruddii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Candidatus Carsonella.  
 OX NCBI\_TaxID=114186;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Thao M.L., Clark M.A., Burckhardt D.H., Moran N.A., Baumann P.;  
 RT "Phylogenetic analysis of vertically transmitted psyllid endosymbionts  
 (Candidatus Carsonella ruddii) based on atpAGD and rpoC; comparisons  
 with 16S-23S rDNA-derived phylogeny.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N DIPHOSPHATE +  
 CC [RNA](N).  
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE  
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1  
 CC BETA' CHAIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.  
 CC EMBL; AF268063; AAK55954.1; -.  
 DR InterPro: IPR001572; RNA\_pol\_B.  
 DR Pfam: PF00562; RNA\_pol\_B; 1.  
 DR DNA-directed RNA polymerase; Transcription; Transferase.  
 KW NON\_TER 1  
 FT SEQUENCE 60 AA; 6617 MW; 9A34D68E7E563350 CRC64;  
 SQ

Query Match 100.0%; Score 20; DB 2; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 5.1e+02;

Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 ||||  
 DB 18 IEGR 21

RESULT 8  
 Q91GA0  
 ID Q91GA0 PRELIMINARY; PRT; 61 AA.  
 AC Q91GA0;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 6.7 kDa protein.  
 CC columbia circovirus.  
 CC Viruses; ssDNA viruses; Circoviridae; Circovirus.  
 OX NCBI\_TaxID=126070;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9030;  
 RA Weston J.H.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9030;  
 RX MEDLINE=21378871; PubMed=11485403;  
 RA Todd D., Weston J.H., Soike D., Smyth J.A.;  
 RT "Genome sequence determinations and analyses of novel circoviruses  
 from goose and pigeon.";  
 RL Virology 286:354-362(2001).  
 DR EMBL; AJ298229; CAC50248.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 61 AA; 6736 MW; 99F8DFF347A1A332 CRC64;

Query Match 100.0%; Score 20; DB 12; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 ||||  
 DB 15 IEGR 18

RESULT 9  
 Q93JW0  
 ID Q93JW0 PRELIMINARY; PRT; 63 AA.  
 AC Q93JW0;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE 4-oxalocrotonate tautomerase.  
 GN TDNL.  
 OS Pseudomonas putida.  
 OG Plasmid pTDNL.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 OX NCBI\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fukumori F.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D85415; BAB62059.1; -.  
 DR InterPro: IPR004370; Taut.  
 DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD019232; Tautomerase; 1.  
 DR TIGRfam; TIGR00013; taut; 1.  
 KW Plasmid.  
 SQ SEQUENCE 63 AA; 6927 MW; 97ED473A49AED63B CRC64;

Query Match 100.0%; Score 20; DB 2; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 9 IEGR 12

## RESULT 10

Q9AP90 PRELIMINARY; PRT; 66 AA.  
AC Q9AP90;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Hypothetical 7.4 kDa protein.  
OS Uncultured bacterium F01.  
OC Bacteria; environmental samples.  
OX NCBI\_TaxID=147490;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSDON-CLASS 7 INTEGRON;  
RX MEDLINE=21100264; PubMed=11166996;  
RA Nield B.S., Holmes A.J., Gillings M.R., Recchia G.D., Mabbutt B.C.,  
RA Nevalainen K.M.H., Stokes H.W.;  
RT "Recovery of new integron classes from environmental DNA.";  
RL FEMS Microbiol. Lett. 195:59-65(2001).  
DR EMBL; AF314190; AAK00306.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 66 AA; 7404 MW; B132766C53690D31 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 66;

Best Local Similarity 100.0%; Pred. No. 6.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 28 IEGR 31

## RESULT 11

O51760 PRELIMINARY; PRT; 66 AA.  
ID O51760  
AC O51760;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein BB0820.  
GN BB0820.  
OS Borrelia burgdorferi (Lyme disease spirochete).  
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
OX NCBI\_TaxID=139;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35210 / B31;  
RX MEDLINE=98065943; PubMed=9403685;  
RA Fraser C.M., Castens S., Huang W.M., Sutton G.G., Clayton R.A.,  
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwin M.,  
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,  
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
RA Smith H.O., Venter J.C.;  
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";  
RL Nature 390:580-586(1997).  
DR EMBL; AE001180; AAC67174.1; -;  
DR TIGR; BB0820; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 66 AA; 7579 MW; 21352F7A53946F45 CRC64;

Query Match 100.0%; Score 20; DB 16; Length 66;

Best Local Similarity 100.0%; Pred. No. 6.5e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 63 IEGR 66

## RESULT 12

O8YGJ9 PRELIMINARY; PRT; 67 AA.  
ID O8YGJ9  
AC O8YGJ9;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein BMEI1159.  
GN BMEI1159.  
OS Brucella melitensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucellaceae; Brucella.  
OX NCBI\_TaxID=29459;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;  
RX MEDLINE=20020109; PubMed=11756688;  
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,  
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,  
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,  
RA Haselkorn R., Kyrpides N., Overbeek R.;  
RT "The genome sequence of the facultative intracellular pathogen Brucella melitensis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
DR EMBL; AB09555; AAL52340.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 67 AA; 7381 MW; 710993A78305185B CRC64;

Query Match 100.0%; Score 20; DB 16; Length 67;

Best Local Similarity 100.0%; Pred. No. 6.6e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 37 IEGR 40

## RESULT 13

O38263 PRELIMINARY; PRT; 68 AA.  
ID O38263  
AC O38263;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Unidentified ORF20.  
OS Lactococcus phage bIL67.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC c2-like viruses.  
OX NCBI\_TaxID=36343;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95111629; PubMed=7812447;  
RA Schouler C., Ehrlich S.D., Chopin M.C.;  
RT "Sequence and organization of the lactococcal prolate-headed bIL67 phage genome.";  
RL Microbiology 140:3061-3069(1994).  
DR EMBL; L33769; AAA74357.1; -;  
SQ SEQUENCE 68 AA; 8153 MW; D096C8C0B41BCF13 CRC64;

Query Match 100.0%; Score 20; DB 9; Length 68;

Best Local Similarity 100.0%; Pred. No. 6.7e+02;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||



Db 27 IEGR 30

## RESULT 14

Q8XRG1 PRELIMINARY: PRT: 70 AA.  
 AC Q8XRG1  
 DT 01-MAR-2002 (TRENBLrel. 20, Created)  
 DT 01-JUN-2002 (TRENBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Probable 4-oxalocrotonate isomerase protein (EC 5.3.2.-).  
 GN RSP0893 OR RS01664.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia  
 OC NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chander M., Choisine N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schliex T.,  
 RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:457-502(2002).  
 DR EMBL; AL646081; CAD18044.1; -;  
 DR InterPro; IPR004370; Taut.  
 DR Pfam; PF01361; Tautomerase; 1.  
 DR ProDom; PD019232; Tautomerase; 1.  
 DR TIGRFAMs; TIGR00013; taut; 1.  
 KW Isomerase; Plasmid; Complete proteome.  
 SQ SEQUENCE 70 AA; 7256 MW; 2D9DED5IA9C4ABCA CRC64;

Query Match 100.0%; Score 20; DB 16; Length 70;  
 Best Local Similarity 100.0%; Pred. No. 6.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 9 IEGR 12

## RESULT 15

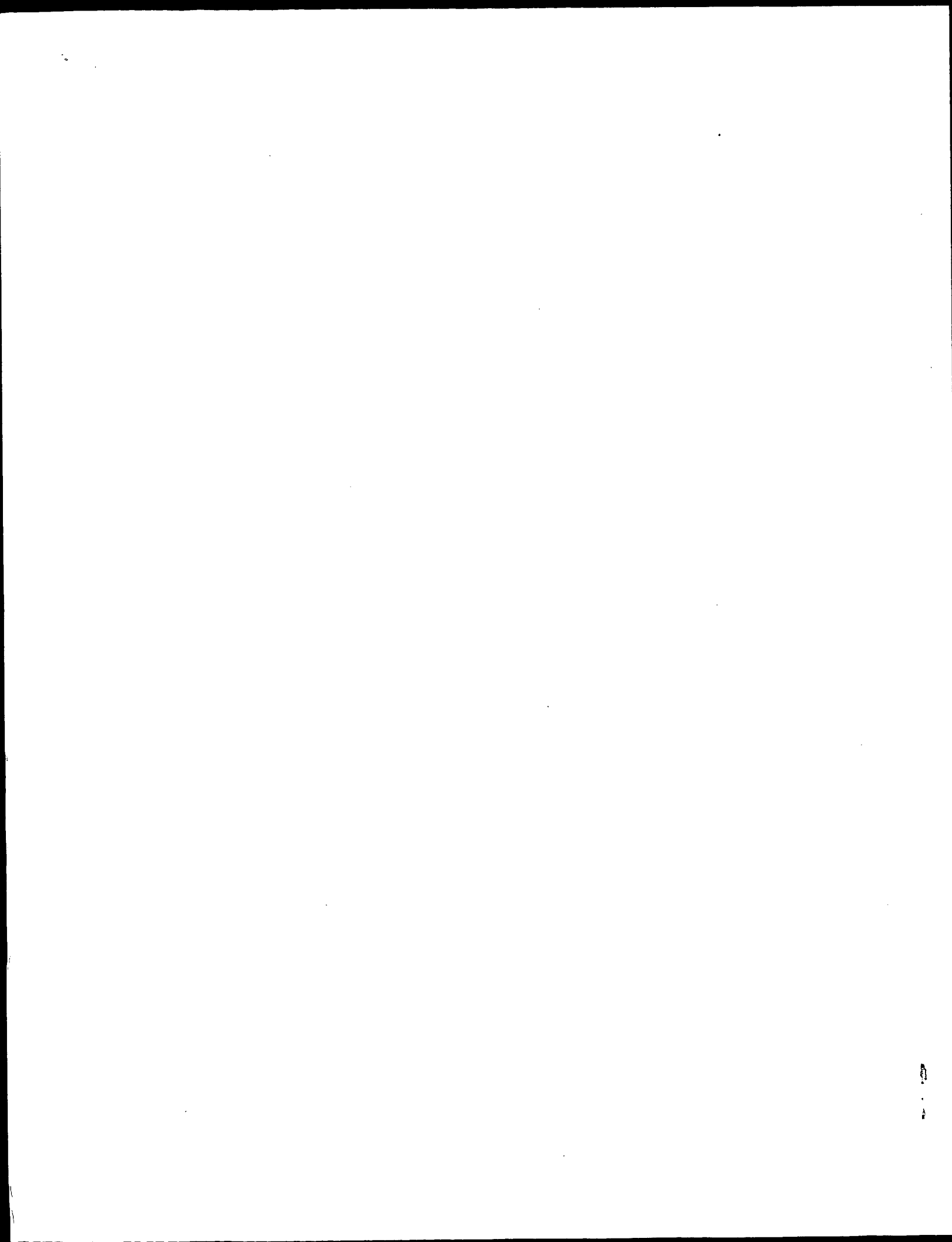
Q9PDH9 PRELIMINARY: PRT: 71 AA.  
 AC Q9PDH9  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Hypothetical protein xfl1400.  
 GN Xfl1400.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 OC NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marcho M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Mirsacca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Pesquero J.B.,  
 RA Quaggio R.B., Pereira G.A.G., Pereira H.A. Jr., de Souza A.A.,  
 RA de Rosa V.E. Jr., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA da Silva A.C.R., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
 RL Nature 406:151-159(2000).  
 DR EMBL; AE003971; AAF84209.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 71 AA; 7647 MW; A85C21C5DBF3FA0E CRC64;

Query Match 100.0%; Score 20; DB 16; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 24 IEGR 27

Search completed: April 13, 2003, 03:22:41  
 Job time : 2.2128 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:36:25 ; Search time 0.0764755 Seconds  
(without alignments)  
2169.392 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result NO.	Query	Score	Match	Length	ID	Description
1	20	100.0	52	1	IF2A_RABIT	P3268 oryctolagus
2	20	100.0	62	1	DMP1_PSESP	P49172 pseudomonas
3	20	100.0	73	1	Y889_ARCFU	O28584 archaeoglob
4	20	100.0	74	1	YAJ9_YEAST	P39549 saccharomyc
5	20	100.0	77	1	ACP_BACHD	O9ka04 bacillus ha
6	20	100.0	86	1	RL27_RICCN	O92990 rickettsia
7	20	100.0	86	1	RL27_RICPR	O92c18 rickettsia
8	20	100.0	105	1	YN77_CAUCR	O9a3s0 caulobacter
9	20	100.0	109	1	BFA3_PSES1	O52440 pseudomonas
10	20	100.0	112	1	PYRE_ENTFA	O07657 enterococcu
11	20	100.0	114	1	Y211_METUA	O60273 methanococc
12	20	100.0	119	1	ACYP_DROME	P56544 drosophila
13	20	100.0	119	1	SZ07_PIG	P43030 sus scrofa
14	20	100.0	129	1	YRDN_BACSU	P45030 bacillus su
15	20	100.0	138	1	V56_HPV49	P36813 human papil
16	20	100.0	146	1	FLAV_TREPA	O38395 treponema p
17	20	100.0	149	1	NORC_PARDE	O51662 paracoccus
18	20	100.0	151	1	PYRI_PYRAE	O8ztg2 pyrobaculum
19	20	100.0	152	1	HSPA_BRAJA	P70917 bradyrhizob
20	20	100.0	153	1	C554_RHOSH	O53142 rhodobacter
21	20	100.0	153	1	CHEW_CAUCR	O87715 caulobacter
22	20	100.0	155	1	CHEW_RHIME	O52881 rhizobium m
23	20	100.0	156	1	GREM_LACIA	O9cht2 lactococcus
24	20	100.0	156	1	YFEB_LACIA	O48660 lactococcus
25	20	100.0	161	1	FABL_CAEEL	O20224 caenorhabdi
26	20	100.0	162	1	GREM_RICPR	P27640 rickettsia
27	20	100.0	163	1	Y22B_HAEIN	O86223 haemophilus
28	20	100.0	166	1	YB38_MYCPN	P75260 mycoplasma
29	20	100.0	175	1	PYRE_HALN1	O9hng2 halobacteri
30	20	100.0	178	1	IF3_BUCAP	P46243 buchnera ap
31	20	100.0	179	1	IF3_BUCAI	P37226 buchnera ap
32	20	100.0	179	1	IF3_PROVU	P3319 proteus vul
33	20	100.0	179	1	SPC4_HUMAN	P21378 homo sapien

34 20 100.0 179 1 SPC4\_MOUSE  
35 20 100.0 179 1 SPC4\_RAT  
36 20 100.0 179 1 SSB\_HELPY  
37 20 100.0 180 1 IF3\_ECOLI  
38 20 100.0 180 1 IF3\_KLEPN  
39 20 100.0 180 1 IF3\_SALTI  
40 20 100.0 180 1 IF3\_SALTI  
41 20 100.0 181 1 KCY\_PYRAB  
42 20 100.0 181 1 SSB\_HELPY  
43 20 100.0 182 1 Y4YS\_RHISN  
44 20 100.0 183 1 HGXR\_TRIFO  
45 20 100.0 183 1 IF3\_SERNA

#### ALIGNMENTS

RESULT 1  
IF2A\_RABIT  
ID IF2A\_RABIT STANDARD; PRT; 52 AA.  
AC P3268;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Eukaryotic translation initiation factor 2 subunit 1 (Eukaryotic  
DE translation initiation factor 2 alpha subunit) (eif-2-alpha) (EIF-  
DE 2alpha) (EIF-2A) (Fragment).  
GN EIF2S1 OR EIF2A.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE, AND PHOSPHORYLATION OF SER-48.  
RC TISSUP-Reticulocytes;  
RX MEDLINE=86304412; PubMed=3745199;  
RA Wettenhall R.E.H., Kudlicki W., Kramer G., Hardesty B.:  
RT "The NH2-terminal sequence of the alpha and gamma subunits of  
RT eukaryotic initiation factor 2 and the phosphorylation site for the  
RT heme-regulated eif-2 alpha kinase.";  
RT J. Biol. Chem. 261:12444-12447(1986).  
RL [2]  
RN SEQUENCE OF 1-23.  
RC TISSUP-Reticulocytes;  
RX MEDLINE=87240215; PubMed=3592677;  
RA Schaefer M.P., Fairwell T., Parker D.S., Knight M., Anderson W.F.,  
RA Safer B.;  
RT "The purification and characterization of subunits alpha, beta, and  
RT gamma from the rabbit reticulocyte eukaryotic initiation factor 2.";  
RL Arch. Biochem. Biophys. 255:337-346(1987).  
CC -1- FUNCTION: Functions in the early steps of protein synthesis by  
CC forming a ternary complex with GTP and initiator tRNA. This  
CC complex binds to a 40S ribosomal subunit, followed by mRNA binding  
CC to form a 43S preinitiation complex. Junction of the 60S ribosomal  
CC subunit to form the 80S initiation complex is preceded by  
CC hydrolysis of the GTP bound to eIF-2 and release of an eIF-2-GDP  
CC binary complex. In order for eIF-2 to recycle and catalyze another  
CC round of initiation, the GDP bound to eIF-2 must exchange with GTP  
CC by way of a reaction catalyzed by eIF-2B.  
CC -1- SUBUNIT: Heterotrimer composed of an alpha, a beta, and a gamma  
CC chain.  
CC -1- PTM: Phosphorylation at least 4 kinases: EIF2AK3/PERK, GCN2, HRI and  
CC PKR. Phosphorylation stabilizes the eIF-2/GDP/eIF-2B complex and  
CC prevents GDP/GTP exchange reaction, thus impairing the recycling  
CC of eIF-2 between successive rounds of initiation and leading to  
CC global inhibition of translation (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE EIF-2-ALPHA FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 SL MOTIF DOMAIN.  
DR InterPro: IPR003029; Sl.  
DR Pfam: PF00575; Sl; 1.  
DR PROSITE: PS0126; Sl; 1.  
DR P3319 proteus vul  
KW Translation regulation; RNA-binding; Phosphorylation.

Q9r0p6 mus musculu  
P42667 rattus norv  
O25841 helicobacte  
P02999 escherichia  
P33318 klebsiella  
Q82613 salmonella  
P33321 salmonella  
Q9uzj6 pyrococcus  
Q9zjy2 helicobacte  
P55727 rhizobium s  
P51900 tritrichomo  
P33320 serratia ma

```
FT DOMAIN 16 >52 S1 MOTIF.
FT MOD_RES 48 48 PHOSPHORYLATION (BY HRI).
FT MOD_RES 51 51 PHOSPHORYLATION (BY EIF2AK3, GCN2, HRI
FT AND PKR) (BY SIMILARITY).
FT UNSURE 10 10
FT UNSURE 44 44
FT UNSURE 52 52
FT CONFLICT 10 10 H -> R (IN REF. 2).
FT CONFLICT 16 16 E -> Q (IN REF. 1).
FT NON_TER 52 52
SQ SEQUENCE 52 AA; 5974 MW; 52E63D8DCEA6B804 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
DB 41 IEGR 44

RESULT 2
DMPL_PSESP STANDARD; PRT; 62 AA.
AC P49172;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 4-oxalotratone tautomerase (EC 5.3.2.-) (4-OT).
GN DMPL.
OS Pseudomonas sp. (strain CF600).
OG Plasmid pVil50.
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92121108; PubMed=1732207;
RA Shingler V., Marklund U., Powlowski J.;
RT "Nucleotide sequence and functional analysis of the complete
RT phenol/3,4-dimethylphenol catabolic pathway of Pseudomonas sp.
RT strain CF600."
RL J. Bacteriol. 174:711-724 (1992).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=96146412; PubMed=8547259;
RA Subramanya H.S., Roper D.I., Dauter Z., Dodson E.J., Davies G.J.,
RA Wilson K.S., Wigley D.B.;
RT "Enzymatic ketonization of 2-hydroxymuconate: specificity and
RT mechanism investigated by the crystal structures of two isomerases."
RL Biochemistry 35:792-802 (1996).
CC -!- FUNCTION: CATALYZES THE KETONIZATION OF 2-HYDROXYMUCONATE
CC STEREOSELECTIVELY TO YIELD 2-OXO-3-HEXENEDIOATE.
CC -!- PATHWAY: META-CLEAVAGE PATHWAY FOR THE DEGRADATION OF TOLUENE,
CC M-XYLENE AND P-XYLENE.
CC -!- SUBUNIT: HOMOHexamER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60835; CAA43229.1; -.
CC PDB; 1OTF; 03-APR-96.
CC InterPro; IPR004370; Taut.
CC Pfam; PF01361; Tautomerase; 1.
CC ProDom; PD019232; Tautomerase; 1.
CC TIGRFAMs; TIGR00013; taut; 1.
CC Isomerase; Plasmid; Aromatic hydrocarbons catabolism; 3D-structure.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 CATALYTIC BASE (BY SIMILARITY).
```

```
SQ SEQUENCE 62 AA; 6974 MW; CEDAB0F92D1E1E0F CRC64;

Query Match 100.0%; Score 20; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
DB 8 IEGR 11

RESULT 3
YG89_ARCFU STANDARD; PRT; 73 AA.
AC O28584;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1689.
GN AF1689.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
CC -!- SIMILARITY: TO M.JANNASCHII MJ122 AND AF1997.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF000987; AAB89570.1; -.
CC TIGR; AF1689; -.
CC InterPro; IPR003847; DUF217.
CC Pfam; PF02697; DUF217; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 73 AA; 8782 MW; D059A3E6B02452C3 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
DB 32 IEGR 35

RESULT 4
YAJ9_YEAST STANDARD; PRT; 74 AA.
ID YAJ9_YEAST
AC P39549;
DT 01-FEB-1995 (Rel. 31, Created)
```

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 8.9 kDa protein in CDC15-YAT1 intergenic region.  
 GN YAR029W OR FUN57.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Bussey H., Keng T., Storms R.K., Vo D.T., Clark M.W., Fortin N.,  
 RA Barton A.B., Kaback D.B., Clark M.W.;  
 RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,  
 RA Storms R.K.;  
 RT "The nucleotide sequence of chromosome I from Saccharomyces  
 RT cerevisiae";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
 CC -1- SIMILARITY: BELONGS TO THE DUF/COS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; L28920; AAC09491.1; -  
 DR SGD; S0000077; YAR029W.  
 DR InterPro: IPR001142; DUF.  
 DR Pfam: PF00674; DUF; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 74 AA; 8910 MW; 11CDFB4F582BBF21 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
 Db 31 IEGR 34

RESULT 5  
 ACP\_BACHD STANDARD; PRT; 77 AA.  
 AC Q9KA04;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE ACP carrier protein (ACP).  
 GN ACPP OR ACPA OR BH2490.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hiram C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty

CC acid biosynthesis (By similarity).  
 CC -1- PATHWAY: De novo fatty acid biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- FM: 4'-phosphopantetheine is transferred from CoA to a specific  
 CC serine of apo-ACP by acpS. This modification is essential for  
 CC activity because fatty acids are bound in thioester linkage to the  
 CC sulfhydryl of the prosthetic group (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AP001515; BAB06209.1; -  
 DR HSP; P80643; LHY8.  
 DR InterPro: IPR003231; Acyl\_carrier.  
 DR InterPro: IPR003880; Pantane\_attach.  
 DR Pfam: PF00550; pp-binding; 1.  
 DR Prodom: PD000887; Acyl\_carrier; 1.  
 DR TIGRFAMs: TIGR00517; acyl\_carrier; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; 1.  
 DR PROSITE; PS00012; ACP\_DOMAIN; 1.  
 KW Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;  
 KW Complete proteome.  
 FT BINDING 37 37 PHOSPHOPANTHEINE (BY SIMILARITY).  
 SQ SEQUENCE 77 AA; 8538 MW; 3575CDFAM5BA15CA CRC64;

Query Match 100.0%; Score 20; DB 1; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
 Db 73 IEGR 76

RESULT 6  
 RL27\_RICCN STANDARD; PRT; 86 AA.  
 ID RL27\_RICCN  
 AC Q92GGO;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 50S ribosomal protein L27.  
 GN RPLA OR RCL163.  
 OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsieae; Rickettsia.  
 OX NCBI\_TaxID=781;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";  
 RL Science 293:2093-2098(2001).  
 CC -1- SIMILARITY: BELONGS TO THE L27P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AE008664; AAL03701.1; -

```

DR InterPro: IPR001684; Ribosomal_L27.
DR Pfam: PF01016; Ribosomal_L27; 1.
DR ProDom: PD003114; Ribosomal_L27; 1.
DR TIGRFAMs: TIGR00062; L27; 1.
DR PROSITE: PS00831; RIBOSOMAL_L27; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 86 AA; 9293 MW; 97ED818F999658C4 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4
Db 63 IEGR 66

RESULT 7
RL27_RICPR STANDARD; PRT; 86 AA.
AC Q9ZCIG;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L27.
DS RPLA OR RP752.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
CC -!- SIMILARITY: BELONGS TO THE L27P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ235273; CAA15180.1; -
CC InterPro: IPR001684; Ribosomal_L27.
CC Pfam: PF01016; Ribosomal_L27; 1.
CC PRINTS: PR00063; RIBOSOMAL_L27.
CC ProDom: PD003114; Ribosomal_L27; 1.
CC TIGRFAMs: TIGR00062; L27; 1.
CC PROSITE: PS00831; RIBOSOMAL_L27; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 86 AA; 9307 MW; 15EF50BD91BE79D8 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4
Db 63 IEGR 66

RESULT 8
YN77_CAUCR STANDARD; PRT; 105 AA.
ID YN77_CAUCR
AC Q9A5S0;

Query Match 100.0%; Score 20; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4
Db 9 IEGR 12

RESULT 9
BPA3_PSES1 STANDARD; PRT; 109 AA.
AC Q52440;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Ferredoxin subunit of biphenyl dioxygenase.
GN BPHA3.
OS Pseudomonas sp. (strain KKS102).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=307;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94324977; PubMed=8048958;
RA Fukuda M., Yasukouchi Y., Kikuchi Y., Nagata Y., Kimbara K.,
RA Horiuchi H., Takagi M., Yano K.;
RT "Identification of the bpha and bphB genes of Pseudomonas sp. strains
RT KKS102 involved in degradation of biphenyl and polychlorinated
RT biphenyls.";
RL Biochem. Biophys. Res. Commun. 202:850-856(1994).
CC -!- FUNCTION: THIS PROTEIN SEEMS TO BE A 2FE-2S FERREDOXIN.
CC -!- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA1 AND BPHA2), A
CC FERREDOXIN (BPHA3) AND A FERREDOXIN REDUCTASE (BPHA4).

```

CC -1- SIMILARITY: BELONGS TO THE BACTERIAL RING-HYDROXYLATING  
 CC DIOXYGENASE FERREDOXIN COMPONENT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D17319; BAA04139.1; -  
 DR HSSP: P37332; LFQT.  
 DR InterPro: IPR001281; Rieske; 1.  
 DR Pfam: PF00355; Rieske; 1.  
 KW Aromatic hydrocarbons catabolism; Electron transport; Iron-sulfur.  
 FT METAL 43 43 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 45 45 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 63 63 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 FT METAL 66 66 IRON-SULFUR (2FE-2S) (POTENTIAL).  
 SQ SEQUENCE 109 AA; 11912 MW; 0D3BBB8C60C01751 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 93 IEGR 96  
 RESULT 10  
 ID PYRE\_ENTFA STANDARD; PRT; 112 AA.  
 AC 007657;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Orotate phosphoribosyltransferase (EC 2.4.2.10) (OPRT) (OPRTase)  
 DE (Fragment).  
 GN PYRE.  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 OG Plasmid pKV48.  
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.  
 OX NCBI\_TaxID=1351;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 47077 / OG1RF;  
 RX MEDLINE-96074317; PubMed-7592480;  
 RA Li X., Weinstock G.M., Murray B.E.;  
 RT "Generation of auxotrophic mutants of Enterococcus faecalis.";  
 RL J. Bacteriol. 177:6866-6873(1995).  
 CC -1- CATALYTIC ACTIVITY: Orotidine 5'-phosphate + diphosphate = orotate  
 CC + 5-phospho-alpha-D-ribose 1-diphosphate.  
 CC -1- PATHWAY: Pyrimidine biosynthesis; fifth step.  
 CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE  
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U24682; AAB61216.1; -  
 DR HSSP: P08870; IOPR.  
 DR InterPro: IPR002375; Pr/py\_rp.transf.  
 DR InterPro: IPR000836; PRtransferase.  
 DR Pfam: PF00156; Pribsyltran; 1.  
 DR PROSITE: PS00103; PUR\_PIR\_PR\_TRANSFER; 1.  
 KW Pyrimidine biosynthesis; Transferase; Glycosyltransferase; Plasmid.

FT NON\_TER 1 1  
 FT ACT\_SITE 75 75 BY SIMILARITY.  
 FT NON\_TER 112 112  
 SQ SEQUENCE 112 AA; 11910 MW; 5B88F43A8D3AF698 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 112;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 83 IEGR 86  
 RESULT 11  
 ID YZ11\_METJA STANDARD; PRT; 114 AA.  
 AC Q60273;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJEC111.  
 GN MJEC111.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE-96337999; PubMed-8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L77118; AAC37084.1; -  
 DR TIGR; MJEC111; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 114 AA; 12826 MW; 99A168C6C180676B CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 77 IEGR 80  
 RESULT 12  
 ID ACYP\_DROME STANDARD; PRT; 119 AA.  
 AC P56544; Q9V3K1;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acylphosphatase (EC 3.6.1.7) (Acylphosphate phosphohydrolase).

GN ACYP OR ACP OR ACPPRO OR CG16870.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98416044; PubMed=9744795;  
 RA Pieri A., Magherini F., Liguori G., Raugeri G., Taddei N.,  
 RA Bozzetti M.P., Cecchi C., Ramponi G.;  
 RT "Drosophila melanogaster acylphosphatase: a common ancestor for  
 RT acylphosphatase isoenzymes of vertebrate species.";  
 RL FEBS Lett. 433:205-210(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=99403001; PubMed=10471707;  
 RA Ashburner M., Misra S., Rote J., Lewis S.E., Blazej R., Davis T.,  
 RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,  
 RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,  
 RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,  
 RA Celniker S., Rubin G.M.;  
 RT "An exploration of the sequence of a 2.9-Mb region of the genome of  
 RT Drosophila melanogaster: the Adh region.";  
 RL Genetics 153:179-219(1999).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Panders R.D.C., Scheeler F., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 CC -!- CATALYTIC ACTIVITY: An acyl phosphate + H(2)O = a fatty acid anion  
 CC + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- MISCELLANEOUS: OPTIMUM PH IS 5.3-6.3.  
 CC -!- SIMILARITY: BELONGS TO THE ACYLPHOSPHATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AJ243543; CAB48386.1; -;  
 CC EMBL: AE003408; AAF44835.1; -;  
 CC EMBL: AE003641; AAF53355.1; -;  
 CC HSSP: P00818; IAPS.  
 CC FlyBase; FBgn0025115; Acyp.  
 CC InterPro; IPR001792; Acylphosphatase.  
 CC Pfam; PF00708; Acylphosphatase; 1.  
 CC PRINTS; PR00112; ACYLPHPTASE.  
 CC ProDom; PD001884; Acylphosphatase; 1.  
 CC PROSITE; PS00150; ACYLPHOSPHATASE\_1; 1.  
 CC PROSITE; PS00151; ACYLPHOSPHATASE\_2; 1.  
 CC Acetylation; Hydrolase.  
 CC INIT\_MET 0 0 BY SIMILARITY.  
 CC MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 CC ACT\_SITE 22 22 POTENTIAL.  
 CC ACT\_SITE 40 40 POTENTIAL.  
 CC FT ACT\_SITE 119 AA; 13566 MW; 803197175FA37795 CRC64;  
 CC SQ SEQUENCE  
 CC Query Match 100.0%; Score 20; DB 1; Length 119;  
 CC Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 IEGR 4  
 CC Db 50 IEGR 53  
 CC  
 CC RESULT 13  
 CC SZ07\_PIG STANDARD; PRT; 119 AA.  
 CC ID SZ07\_PIG  
 CC AC P43030;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Platelet basic protein precursor (PBP) (Small inducible cytokine B7)  
 CC DE (CXCL7).  
 CC GN SCYB7 OR PPBP.  
 CC OS Sus scrofa (Pig).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 CC OX NCBI\_TaxID=9823;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A., AND SEQUENCE OF 40-119.  
 CC RC TISSUE=Platelet;  
 CC RX MEDLINE=94229068; PubMed=7513641;  
 CC RA Power C.A., Proudfoot A.E.I., Magnat E., Bacon K.B., Wells T.N.C.;  
 CC RT "Molecular cloning and characterisation of a neutrophil chemotactic  
 CC protein from porcine platelets.";  
 CC RL Eur. J. Biochem. 221:713-719(1994).  
 CC CC -!- FUNCTION: CHEMOATTRACTANT FACTOR FOR NEUTROPHILS.  
 CC CC -!- MASS SPECTROMETRY: MW=8597.5; METHOD=Electrospray; RANGE=40-119.  
 CC CC -!- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE  
 CC C-X-C) (CHEMOKINE CXC).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X77935; CAA54907.1; -;  
 CC HSSP: P02775; INAP.  
 CC DR



DR InterPro; IPR001089; CXC\_chemkine.smll.  
 DR InterPro; IPR001811; Chemokine\_IL8.  
 DR Pfam; PF00048; IL8; 1.  
 DR PRINTS; PR00437; SMALLCYTKCXC.  
 DR SMART; SM00199; SCY; 1.  
 DR PROSITE; PS00471; SMALL\_CYTOKINES\_CXC; 1.  
 KW Cytokine; Growth factor; Chemotaxis; Mitogen; Platelet; Signal.  
 FT SIGNAL 1 33 PROBABLE.  
 FT PROPEP 34 39  
 FT CHAIN 40 119 PLATELET BASIC PROTEIN.  
 FT DISULFID 54 80 BY SIMILARITY.  
 FT DISULFID 56 96 BY SIMILARITY.  
 SQ SEQUENCE 119 AA; 12615 MW; 79E1D409CDD06B32 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 DB 43 IEGR 46

RESULT 14  
 YRDN\_BACSU STANDARD; PRT; 129 AA.  
 AC P94502; O08187;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein yrdn.  
 GN YRDN.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97175526; PubMed=9023181;  
 RX MEDLINE=97175526; PubMed=9023181;  
 RA Belitsky B.R., Sonenshein A.L.;  
 RT "Altered transcription activation specificity of a mutant form of  
 Bacillus subtilis GltR, a LysR family member.";  
 RL J. Bacteriol. 179:1035-1043(1997).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=97453479; PubMed=9308178;  
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,  
 RA Dueterhoeft A., Ehrlich S.D.;  
 RT "Sequence of the Bacillus subtilis genome region in the vicinity of  
 the lev operon reveals two new extracytoplasmic function RNA  
 polymerase sigma factors SigV and SigZ.";  
 RL Microbiology 143:2939-2943(1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dueterhoeft A., Ehrlich S.D., Emmerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Coffeau A., Golightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Halbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Henaut A.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapide A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzensegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunaga K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis.";  
 RT Nature 390:249-256(1997).  
 RL Nature 390:249-256(1997).  
 CC -!- FUNCTION: PUTATIVE TARGET OF GLTR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U79494; AAB47964.1; -;  
 DR EMBL; U79494; AAB47964.1; -;  
 DR EMBL; U93876; AAB80906.1; -;  
 DR EMBL; U93876; AAB80906.1; -;  
 DR Subtilist; BG12287; yrdn.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 129 AA; 14682 MW; 54DDEC8184E7461F CRC64;

Query Match 100.0%; Score 20; DB 1; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 DB 9 IEGR 12

RESULT 15  
 VE6\_HPV49  
 ID VE6\_HPV49 STANDARD; PRT; 138 AA.  
 AC P36813;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE E6 protein.  
 DE E6.  
 GN Human papillomavirus type 49.  
 OS Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=10616;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94265501; PubMed=8205838;  
 RA Delius H., Hofmann B.;  
 RT "Primer-directed sequencing of human papillomavirus types.";  
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).  
 CC -!- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE  
 CC STRANDED DNA (IN VITRO).  
 CC -!- SUBCELLULAR LOCATION: Nuclear matrix-associated.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; X74480; CAA52579.1; -;

DR PIR; S36567; S36567.  
 DR InterPro; IPR001334; E6.  
 DR Pfam; PF00518; E6; 1.  
 KW Early protein; DNA-binding; Nuclear protein; Zinc-finger.  
 FT ZN\_FING 25 61  
 FT ZN\_FING 98 134  
 SQ SEQUENCE 138 AA; 16202 MW; 69AD429D88C08ADB CRC64;

Query Match 100.0%; Score 20; DB 1; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
 ||||  
 Db 83 IEGR 86

Search completed: April 13, 2003, 03:21:34  
 Job time : 1.07648 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:34:46 ; Search time 0.222776 seconds  
(without alignments)  
2392.545 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	4	5 AAP40354	Sequence of hydrol
2	20	100.0	4	6 AAP50015	Sequence of a clea
3	20	100.0	4	12 AAR13804	Factor Xa substrat
4	20	100.0	4	12 AAR12870	Chromogenic substr
5	20	100.0	4	13 AAR27101	Sequence of tetrap
6	20	100.0	4	14 AAR32109	Sequence of blood
7	20	100.0	4	14 AAR39390	Factor Xa cleavage
8	20	100.0	4	15 AAR54863	Factor Xa cleavage
9	20	100.0	4	15 AAR48071	Biological protect
10	20	100.0	4	15 AAR60504	Factor Xa cleavage

11	20	100.0	4	15 AAR60693	H. halobium bacter
12	20	100.0	4	16 AAR66901	Blood coagulation
13	20	100.0	4	16 AAR8219	Lactoferrisin anti
14	20	100.0	4	16 AAR83116	Factor Xa recognit
15	20	100.0	4	16 AAR73943	Endotoxin assay pe
16	20	100.0	4	17 AAR88138	Processing site fo
17	20	100.0	4	18 AAW23461	Factor Xa protease
18	20	100.0	4	18 AAW32850	Protecting group w
19	20	100.0	4	18 AAW18874	Peptide for diagno
20	20	100.0	4	18 AAW21759	Factor Xa cleavage
21	20	100.0	4	19 AAW70275	Restriction protea
22	20	100.0	4	20 AAY30775	Peptide used to de
23	20	100.0	4	20 AAY33403	Factor Xa protease
24	20	100.0	4	20 AAY41678	Factor Xa substrat
25	20	100.0	4	20 AAY29847	Peptide Seq ID No:
26	20	100.0	4	20 AAY31224	Fxa excision site
27	20	100.0	4	20 AAY28949	Peptide adjacent t
28	20	100.0	4	20 AAY06472	Factor Xa recognit
29	20	100.0	4	21 ABB05354	Angiotensin inhibit
30	20	100.0	4	21 AAB37322	Peptide linker #7
31	20	100.0	4	21 AAB15679	Factor Xa-sensitiv
32	20	100.0	4	21 AAB22832	Factor Xa cleavage
33	20	100.0	4	21 AAB26264	Factor Xa cleavage
34	20	100.0	4	21 AAB15083	Factor Xa cleavage
35	20	100.0	4	21 AAB20773	Factor Xa enzyme s
36	20	100.0	4	21 AAY96957	Factor Xa recognit
37	20	100.0	4	21 AAY90829	Factor Xa peptide
38	20	100.0	4	21 AAY90453	Factor Xa substrat
39	20	100.0	4	21 AAY76807	Factor Xa substrate
40	20	100.0	4	21 AAY44371	Factor Xa recognit
41	20	100.0	4	21 AAY57591	Factor Xa recognit
42	20	100.0	4	22 AAO17349	Soluble fibrin ana
43	20	100.0	4	22 AAU11213	Factor Xa protease
44	20	100.0	4	22 AAE13083	Factor Xa site-spe
45	20	100.0	4	22 AAG67095	Peptide recognised

#### ALIGNMENTS

##### RESULT 1

AAP40354

ID AAP40354 standard; peptide; 4 AA.

XX AAP40354;

XX 03-OCT-2002 (updated)  
DT 11-FEB-1992 (first entry)

XX Sequence of hydrolysable naphthalene derivs. used as substrate in an assay to determine the activity of Factor Xa.

XX Naphthalene derivative; enzyme assay.

XX Synthetic.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1

FT Modified-site 4 /label= benzol-ile

FT Modified-site 4 /label= Arg-alpha-naphthyl ester

XX DE3327873-A.

XX 09-FEB-1984.

XX 02-AUG-1983; 83DE-3327873.

XX 03-AUG-1982; 82JP-0135534.

XX (TORI ) TORII & CO LTD.

PI Fujii S, Sugiyama S, Sawai S;

XX WPI; 1984-038201/07.  
XX  
XX Determn. of enzyme, inhibitor, activator and zymogen activity -  
PT by reacting the enzyme with hydrolysable naphthalene derivs.  
PT which is then allowed to form pigment with fast red-ITR-salt  
XX  
XX  
XX Example; Page 21; 42pp; German.  
XX  
XX The claimed method for the determn. of enzyme, inhibitor, activator  
CC and zymogen activity involves reacting the enzyme with hydrolysable  
CC naphthalene derivs. which is then allowed to form pigment with fast  
CC red-ITR-salt. The process can be used in quality control of enzyme  
CC preps., in clinical investigations and in the diagnosis of various  
CC illnesses which depend on the enzyme content of blood or urine.  
CC (Updated on 03-OCT-2002 to add missing OS field.)  
XX  
XX SQ Sequence 4 AA;  
XX  
XX Query Match . 100.0%; Score 20; DB 5; Length 4;  
XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 IEGR 4  
XX  
XX Db 1 IEGR 4  
XX  
XX  
XX RESULT 2  
XX AAP50015  
XX ID AAP50015 standard; Protein; 4 AA.  
XX  
XX AC AAP50015;  
XX  
XX DT 16-AUG-2002 (updated)  
XX DT 09-SEP-1991 (first entry)  
XX  
XX Sequence of a cleavage site which is specifically cleaved by  
DE blood coagulation Factor Xa.  
XX  
XX Fusion protein; cleavage site; proteolytic cleavage.  
XX  
XX Homo sapiens.  
XX Bos taurus.  
XX  
XX PN EP161937-A.  
XX  
XX PD 21-OCT-1985.  
XX  
XX PF 15-MAY-1985; 85EP-0303414.  
XX  
XX PR 16-MAY-1984; 84GB-0012517.  
XX PR 15-MAY-1985; 85GB-0012333.  
XX  
XX (CELL-) CELTECH LTD.  
XX  
XX Nagai K, Thogersen HC;  
XX  
XX WPI; 1985-291163/47.  
XX N-PSDB; AAN50051, AAN50052.  
XX  
XX DNA sequence coding for cleavage site - which is specifically  
PT cleaved by blood coagulation factor XA  
XX  
XX Claim 4; Page 23; 29pp; English.  
XX  
XX Vectors comprising the SQs in AAN50051 and AAN50052 are also claimed, as  
CC is a process for the prodn. of a protein or peptide prod. in native  
CC form which uses the vectors to produce the protein or peptide as a  
CC fusion protein which is then cleaved with Factor Xa.  
CC (Updated on 16-AUG-2002 to add missing OS field.)  
XX  
XX SQ Sequence 4 AA;  
XX

XX Query Match 100.0%; Score 20; DB 6; Length 4;  
XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 IEGR 4  
XX  
XX Db 1 IEGR 4  
XX  
XX  
XX RESULT 3  
XX AAR13804  
XX ID AAR13804 standard; Protein; 4 AA.  
XX  
XX AC AAR13804;  
XX  
XX DT 07-NOV-1991 (firs entry)  
XX  
XX DE Factor Xa substrate peptide (2).  
XX  
XX KW Assay; factor Xa; substrate; affinity.  
XX  
XX OS Synthetic.  
XX  
XX FH Key Location/Qualifiers  
XX FT Modified-site 1  
XX FT Modified-site 2 /note= "methoxymalonyl-D-isoleucine"  
XX FT Modified-site 4 /note= "L-Glu(OMe)"  
XX FT Modified-site 4 /note= "L-Arg-p-nitroanilino"  
XX  
XX PN W09112338-A.  
XX  
XX PD 22-AUG-1991.  
XX  
XX PF 31-DEC-1990; 90WO-FR00974.  
XX  
XX PR 19-FEB-1990; 90FR-0001965.  
XX  
XX (SERB-) SERBIO.  
XX  
XX Quentin G, Martinoli JL;  
XX  
XX WPI; 1991-267149/36.  
XX  
XX New labelled tri: or tetra:peptide derivs. - substrates for  
PT factor Xa assay, with better affinity, selectivity or water  
PT solubility  
XX  
XX Claim 8 (j); Page 40; 46pp; French.  
XX  
XX The peptides represented in AAR13803-12 are examples of a generic  
CC formula. They are substrates for assaying factor Xa (an enzyme  
CC involved in haemostasis). Compared with known substrates they have  
CC better affinity, selectivity and/or water solubility.  
XX  
XX SQ Sequence 4 AA;  
XX  
XX Query Match 100.0%; Score 20; DB 12; Length 4;  
XX Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
XX Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 IEGR 4  
XX  
XX Db 1 IEGR 4  
XX  
XX  
XX RESULT 4  
XX AAR12870  
XX ID AAR12870 standard; Protein; 4 AA.  
XX  
XX AC AAR12870;  
XX

XX 24-SEP-1991 (first entry)  
 XX Chromogenic substrate S-2860.  
 DE Bacterial endotoxin; Limulus amoebocyte clotting enzyme.  
 XX Synthetic.  
 KW Key Location/Qualifier  
 XX Modified-site 3  
 FT /label= glycolic acid  
 FT W09109052-A.  
 XX 27-JUN-1991.  
 PD 03-DEC-1990; 90WO-SE00797.  
 XX 12-DEC-1989; 89SE-0004188.  
 XX (KABI ) KABIVITRUM AB.  
 PA Arielly S;  
 PI WPI; 1991-208087/28.  
 DR New peptide derivs. - used as substrates for quantitative  
 PT determ. of bacterial endotoxins in fluids, foods or  
 PT pharmaceuticals.  
 XX Claim 3; Page 15; 21pp; English.  
 XX The N-terminal has an alpha-Ac gp., and the C-terminal has a  
 CC 4-nitroaniline (pNA) gp. The peptide is a specific example of the  
 CC generic formula: R1-A1-A2-A3-A4-R2  
 CC R1-H or a protective gp;  
 CC A1-H, Ile, Leu, or Val;  
 CC A2=Glu, Asp, Ser, or Thr;  
 CC A3=Gly or glycolic acid (Glyc);  
 CC A4=Arg or Lys; and  
 CC R2=4-nitro aniline  
 CC with the proviso that A3=Gly when A4=Lys and A3=Glyc when A4=Arg.  
 CC The peptides are used as substrates for the quantitative determ.  
 CC of bacterial endotoxins e.g. in physiological fluids, food or  
 CC pharmaceuticals. They have higher activity than the known substrate  
 CC S-2423 which has Gly-Arg as the C-terminal sequence. They are  
 CC prep. by solid phase synthesis.  
 CC See also AAR12867-R12869.  
 XX Sequence 4 AA;  
 SQ Query Match 100.0%; Score 20; DB 12; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 DB 1 IEGR 4  
 RESULT 5  
 AAR27101  
 ID AAR27101 standard; peptide; 4 AA.  
 XX AAR27101;  
 AC AAR27101;  
 XX 03-FEB-1993 (first entry)  
 DT Sequence of tetrapeptide recognised by factor Xa.  
 DE Factor Xa; enzymatic hydrolysis; soluble precursor.  
 XX

XX Synthetic.  
 OS EP505921-A.  
 XX 30-SEP-1992.  
 XX 19-MAR-1992; 92EP-0104753.  
 XX 28-MAR-1991; 91IT-0000861.  
 XX (ENIE ) ENIRICRCH SPA.  
 XX Galli G, Grandi G;  
 XX WPI; 1992-325240/40.  
 XX Mature polypeptide prodn. - by cleavage of soluble precursor using  
 PT immobilised factor Xa (Eng)  
 PT Claim 1; Page 5; 9pp; English.  
 PS The inventors claims a method for the preparation of a mature  
 XX polypeptide which involves the enzymatic cleavage of a soluble  
 CC precursor of the polypeptide using factor Xa. In particular the  
 CC method is used for the preparation of mature human growth hormone.  
 CC Eg, Factor Xa is immobilised on a solid insoluble support which is  
 CC brought into contact with a soluble precursor of the polypeptide,  
 CC hGH, which has the formula AAR27102-R27101-hGH. AAR27102 confers  
 CC solubility on hGH; and AAR27101 is a tetrapeptide recognised by factor  
 CC Xa.  
 XX Sequence 4 AA;  
 SQ Query Match 100.0%; Score 20; DB 13; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 DB 1 IEGR 4  
 RESULT 6  
 AAR32109  
 ID AAR32109 standard; Protein; 4 AA.  
 XX AAR32109;  
 AC AAR32109;  
 XX 11-JUL-1993 (first entry)  
 DT Sequence of blood coagulation factor Xa recognition site.  
 XX Factor Xa linker; fusion gene; blood coagulation factor.  
 XX Synthetic.  
 OS EP532043-A.  
 XX 17-MAR-1993.  
 XX 11-SEP-1992; 92EP-0115607.  
 XX 13-SEP-1991; 91JP-0234430.  
 XX (HITA ) HITACHI LTD.  
 XX Harada Y, Sakamoto T, Senda T, Takamoto K;  
 XX WPI; 1993-087069/11.  
 XX Factor Xa linker DNA - used to construct fusion genes which can  
 PT be expressed in large quantities in E coli for prodn.of foreign

## PT proteins

PS Disclosure; column 1; 12pp; English.

CC The DNA in AAQ38417 and AAQ38418 encodes blood coagulation factor Xa recognition sequence IEGR (AAR32109). GA is added immediately after the CDS to form restriction enzyme NruI recognition site. A DNA CC having inverted repeats of the first 11 nucleotides of the CDS is CC ligated to form a palindrome sequence of 22 nucleotide pairs.

XX Sequence 4 AA;

Query Match 100.0%; Score 20; DB 14; Length 4;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

## RESULT 7

AAAR39390  
ID AAR39390 standard; Protein; 4 AA.

XX AAR39390;

XX 12-JAN-1994 (first entry)

DE Factor Xa cleavage site.

KW Rapid; protein isolation; calcium; presence; fusion protein;  
KW affinity chromatography.

XX Synthetic.

PH Key Location/Qualifiers

FT Cleavage-site 1..4  
/note= "factor Xa cleavage site"

XX WO9313211-A.

XX 08-JUL-1993.

XX 29-DEC-1992; 92WO-US11270.

XX 03-JAN-1992; 92US-0816679.

XX (OKLA-) OKLAHOMA MED RES FOUND.

XX Esmon CT, Morrissey JH, Rezale A;

XX WPI; 1993-227327/28.

XX Fusion protein allowing rapid isolation of protein from soln. -  
PT comprises desired protein and epitope recognised by monoclonal  
PT antibody HPC-4

XX Example; Page 30; 42pp; English.

XX The sequence is that of the factor Xa cleavage site which was used  
CC in the prepn. of a fusion protein comprising the epitope for the  
CC Ca++ dependent monoclonal antibody HPC-4 and a protein to be  
CC isolated. This fusion protein may be isolated using HPC-4-based  
CC affinity chromatography, the protein can be recovered in a single  
CC chromatographic step using immobilised HPC-4, and released from the  
CC fusion protein by cleavage with factor Xa.

XX Sequence 4 AA;

Query Match 100.0%; Score 20; DB 14; Length 4;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

## RESULT 8

AAAR54863

ID AAR54863 standard; peptide; 4 AA.

XX AAR54863;

XX 09-JAN-1995 (first entry)

DE Factor Xa cleavage site.

KW Human calcitonin; recombinant production; fusion protein;  
KW Chloramphenicol acetyl transferase; Factor Xa cleavage site;  
KW cell-free protein synthesis system.

XX Homo sapiens.

PH Key Location/Qualifiers

FT Cleavage-site 1..4  
/label= Factor\_Xa\_recognition\_site  
/note= "fused to N-terminus of hCT"

XX JP06098790-A.

XX 12-APR-1994.

XX 04-MAR-1992; 92JP-0081480.

XX 04-MAR-1992; 92JP-0081480.

XX (KOBM ) KOBE STEEL LTD.

XX WPI; 1994-155930/19.

XX Prodn. of polypeptide in non-cellular protein synthesis system -  
PT by constructing a fusion gene with chloramphenicol acetyl  
PT transferase gene and opt. cleaving the polypeptide from the  
PT resultant fusion protein.

XX Example 1; Page 4; 7pp; Japanese.

XX A synthetic coding sequence for human calcitonin was constructed  
CC which coded for hCT (1-32) with an additional amidated C-terminal  
CC Gly residue. The coding sequence was fused immediately downstream of  
CC a region coding for a Factor Xa cleavage site (Ile-Glu-Gly-Arg),  
CC which was itself in-frame, within the chloramphenicol acetyl  
CC transferase (CAT) gene and downstream of a lac or tac promoter.  
CC Human CT can be recovered from the resulting CAT-(IEGR)-hCT fusion  
CC protein by cleavage with factor Xa. The synthetic hCT could be  
CC replaced by sequences coding for other useful polypeptides.

XX Sequence 4 AA;

Query Match 100.0%; Score 20; DB 15; Length 4;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

## RESULT 9

AAAR48071

ID AAR48071 standard; peptide; 4 AA.

XX AAR48071;

XX AAR48071;

DT 13-JUL-1994 (first entry)  
 XX Biological protecting gp. contg. Factor Xa cleavage site.  
 XX  
 XX C-terminal alpha-amide polypeptide; amidation; protecting group;  
 KW N-terminal alpha-acetyl polypeptide; acetylation;  
 KW recombinant multicopy fusion protein; interconnecting peptide;  
 KW intraconnecting peptide; Blood coagulation Factor Xa;  
 KW recognition site.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Cleavage-site 1..4  
 FT /label= Factor\_Xa\_recognition\_site  
 XX  
 XX WO9401451-A.  
 PN  
 XX 20-JAN-1994.  
 PD  
 XX 13-JUL-1993; 93WO-US06591.  
 PF  
 XX 13-JUL-1992; 92US-0912798.  
 PR  
 XX (BION-) BIONEERASKA INC.  
 PA  
 XX Coolidge TR, Holmquist B, Stout J, Wagner FW;  
 PI WPI; 1994-034983/04.  
 XX N-PSDB; AAQ55445.  
 DR  
 XX Terminal modification of recombinant single copy polypeptide - by  
 PT protecting, modifying and de-protecting polypeptide, e.g. fusion  
 PT protein, contg. biologically added protecting gp.  
 XX  
 XX Disclosure and Claim 10; Page 22; 102pp; English.  
 PS  
 XX In order to produce recombinant polypeptides with C-terminal  
 CC alpha-amide groups and/or N-terminal acetyl groups, fusion proteins  
 CC are formed in which the relevant terminus (termini) of the  
 CC polypeptide is protected by a biological protecting group. The  
 CC protecting group is a peptide or amino acid having at least one  
 CC cleavage site for its removal. The sequence AAR48071 represents a  
 CC suitable biological protecting group, i.e. a Factor Xa cleavage  
 CC site.  
 XX  
 XX Sequence 4 AA;  
 SQ  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db  
 1 IEGR 4  
 RESULT 10  
 AAR60504  
 ID AAR60504 standard; Peptide; 4 AA.  
 XX  
 XX AAR60504;  
 AC  
 XX 22-MAR-1995 (first entry)  
 DT  
 XX Factor Xa cleavage site.  
 DE  
 XX Serine protease; Factor-Xa; recognition site;  
 KW fusion protein cleavage; protein folding.  
 KW  
 XX Synthetic.  
 OS  
 XX WO9418227-A.  
 PN

XX 18-AUG-1994.  
 PD  
 XX 04-FEB-1994; 94WO-DK000054.  
 PF  
 XX 04-FEB-1993; 93DK-0000130.  
 PR 05-FEB-1993; 93DK-0000139.  
 PR 03-DEC-1993; 93WO-GB02492.  
 XX (DENZ-) DENZYME APS.  
 FA  
 XX Etzerodt M, Holtet TL, Thogersen HC;  
 PI WPI; 1994-279681/34.  
 DR  
 XX Refolding of polypeptide molecules - using a cyclic process  
 PT involving denaturing and renaturing conditions to produce a  
 PT correctly folded prod  
 PT  
 XX Disclosure; Page 126; 202pp; English.  
 PS  
 XX Novel recognition sites for cattle Factor-Xa are given in  
 CC AAR60504 and AAR60506-12. Such sequences may replace the Factor-Xa  
 CC recognition sequence (AAR60503) in recombinant fusion proteins to  
 CC facilitate cleavage by a new serine protease (given in AAR60502).  
 XX  
 XX Sequence 4 AA;  
 SQ  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 1 IEGR 4  
 RESULT 11  
 AAR60693  
 ID AAR60693 standard; Protein; 4 AA.  
 XX  
 XX AAR60693;  
 AC  
 XX 25-JUN-1995 (first entry)  
 DT  
 XX H. halobium bacteriorhodopsin protease cleavage site.  
 DE  
 XX Halobacteria; expression.  
 KW  
 XX Halobacterium halobium strain R1.  
 OS  
 XX WO9421789-A.  
 PN  
 XX 29-SEP-1994.  
 PD  
 XX 28-FEB-1994; 94WO-US02388.  
 PF  
 XX 25-MAR-1993; 93US-0038662.  
 PR  
 XX (REGC ) UNIV CALIFORNIA.  
 PA  
 XX Betlach MC, Turner GJ;  
 PI WPI; 1994-317010/39.  
 DR  
 XX Expression of heterologous proteins in halo-bacteria - using  
 PT regulatory and stop sequences from halo-bacteria, pref. the  
 PT bacterio-rhodopsin gene.  
 PT  
 XX Disclosure; Page 31; 118pp; English.  
 PS  
 XX The sequence is that of the Halobacterium halobium bacteriorhodopsin  
 CC protease cleavage site. It is used to exemplify a new expression  
 CC

CC vector for producing heterologous polypeptides in a halobacterial host.  
 CC See also AAR60691-9.

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

RESULT 12

AAR86901  
 ID AAR86901 standard; Peptide; 4 AA.

XX AC AAR86901;

XX DT 16-MAY-1996 (first entry)

XX DE Blood coagulation factor Xa selective cleavage site.

XX KW GRB; growth factor receptor bound; tyrosine kinase; regulation;  
 cell growth; cellular metabolism; screening; signal transduction;  
 KW cancer; diabetes; CORT technique; cloning of receptor targets.

XX OS Homo sapiens.

XX PN WO9524426-A1.

XX PD 14-SEP-1995.

XX PF 13-MAR-1995; 95WO-US03385.

XX PR 11-MAR-1994; 94US-0208887.

XX PA (UUNY ) UNIV NEW YORK STATE.

XX PI Margolis BL, Schlessinger J, Skolnik EY;

XX DR WPI; 1995-328235/42.

XX PT DNA encoding tyrosine kinase-binding proteins - used to screen  
 agents capable of modulating cell growth or cellular metabolism

XX PS Disclosure; Page 46; 215pp; English.

XX CC AAR86901 and AAR86902 are selective cleavage sites which can be  
 incorporated into a plasmid used for the recombinant production of GRB  
 proteins. Using a new cloning technique, CORT (cloning of receptor  
 targets) several new TK binding proteins were isolated. Growth factor  
 receptor bound proteins GRB-1, GRB-2, GRB-3, GRB-4, GRB-7 and GRB-10 were  
 isolated using this method. The proteins bind to a tyrosine-  
 phosphorylated domain of a eukaryotic TK. GRB proteins can be used for  
 screening agents which are capable of modulating cell growth that  
 occurs via signal transduction through TKs. Such agents can be used to  
 prevent or inhibit cell growth or to counteract tumour development.  
 CC GRB proteins are also useful for identifying susceptibility to  
 diseases associated with alterations in cellular metabolism mediated by  
 CC TK pathways e.g. cancer and diabetes.

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

RESULT 13

AAR88219  
 ID AAR88219 standard; peptide; 4 AA.

XX AC AAR88219;

XX DT 13-JUN-1996 (first entry)

XX DE Lactoferrisin antibacterial peptide associated protease fragment.

XX KW Recombinant vector; generic; lactoferrisin; antibacterial peptide;  
 regulatory sequence; tac; promoter; shuttle; vector; pGEX2; GAL1;  
 KW pKOM2; Rous Sarcoma Virus; long terminal repeat; pRSVNot;  
 KW lactoferricin; protease fragment.

XX OS Synthetic.

XX PN JP07274970-A.

XX PD 24-OCT-1995.

XX PF 01-APR-1994; 94JP-0085244.

XX PR 01-APR-1994; 94JP-0085244.

XX PA (MORG ) MORINAGA MILK IND CO LTD.

XX DR WPI; 1995-399338/51.

XX PT Recombinant vector contg. lactoferrisin gene - used to prepare an  
 antibacterial peptide

XX PS Example 1; Page 12; 18pp; Japanese.

XX CC A recombinant vector in which a DNA sequence encoding at least the  
 generic lactoferrisin antibacterial peptide AAR88216 is inserted,  
 CC pref. downstream of the vector's regulatory sequence, is claimed.  
 CC The protease fragment AAR88219 and the DNA sequences AAT08774-79 were  
 used in the construction of such a vector, where the regulatory  
 CC sequence is the tac promoter from shuttle vector pGEX2, the GAL1  
 CC promoter from vector pKOM2 or Rous Sarcoma Virus long terminal  
 CC repeat from vector pRSVNot.

SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

RESULT 14

AAR83116  
 ID AAR83116 standard; Peptide; 4 AA.

XX AC AAR83116;

XX DT 29-APR-1996 (first entry)

XX DE Factor-Xa recognition sequence.

XX KW Calcitonin; transgenic animal; milk; Paget disease; hypercalcaemia;  
 osteoporosis; fusion protein cleavage; Factor-Xa.

XX OS Synthetic.

XX PN WO9527782-A1.

XX PD 19-OCT-1995.



XX 05-APR-1995; 95WO-GB00769.  
 XX  
 XX 08-APR-1994; 94GB-0006974.  
 XX  
 XX (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.  
 XX  
 XX Cottingham IR, Garner I;  
 XX  
 XX WPI; 1995-366387/47.  
 XX  
 XX Peptide, pref. calcitonin, prodn. in transgenic non-human mammal -  
 PT by expressing fusion protein in the milk followed by cleavage and  
 PT sepn., used to treat Paget's disease, hypercalcaemic shock etc.  
 XX  
 XX Claim 13; Page 21; 26pp; English.  
 XX  
 XX The recognition sequence (AAR83116) for Factor-Xa may be included in  
 CC the linker segment of a fusion protein, comprising e.g. calcitonin  
 CC and alpha-lactalbumin, produced in the milk of a transgenic mammal.  
 CC Cleavage of the fusion protein with Factor-Xa yields alpha-  
 CC lactalbumin, which is removed by an affinity method, and purified  
 CC calcitonin of therapeutic appin. Alternatively, the linker  
 CC comprises the activation peptide (AAR83118) of trypsinogen, which  
 CC includes the recognition sequence (AAR83117) for enterokinase.  
 XX  
 XX Sequence 4 AA;  
 SQ

Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 1 IEGR 4

RESULT 15  
 AAR73943  
 ID AAR73943 standard; peptide; 4 AA.  
 XX  
 XX AAR73943;  
 XX  
 XX 24-OCT-1995 (first entry)  
 XX  
 XX Endotoxin assay peptide.  
 XX  
 XX Endotoxin; assay; Limulus amoebocyte lysate; cascade reaction.  
 XX  
 XX Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Modified-site 1 /note= "N-terminal Boc group"  
 FT Modified-site 4 /note= "C-terminal pNa group"  
 FT  
 XX EP649021-A.  
 XX  
 XX 19-APR-1995.  
 XX  
 XX 30-SEP-1994; 94EP-0115442.  
 XX  
 XX 30-SEP-1993; 93JP-0265479.  
 XX  
 XX (SEKG ) SEIKAGAKU KOGYO CO LTD.  
 XX  
 XX Oda T, Tamura H, Tanaka S;  
 XX  
 XX WPI; 1995-148858/20.  
 XX  
 XX Limulus amoebocyte lysate reagent for endotoxin determ. - contg.  
 PT alkyl glucoside to inhibit activation of factor G by

PT (1-3)-beta-D-glucan  
 XX  
 XX Disclosure; Page 15; 20pp; English.  
 XX  
 XX The peptides given in AAR73942-44 are used in endotoxin-specific  
 CC assays in which clotting enzymes, formed in cascade reactions  
 CC initiated by addition of endotoxin, hydrolyze an amide bond in  
 CC the peptides, thereby liberating chromogenic p-nitroaniline.  
 XX  
 XX Sequence 4 AA;  
 SQ

Query Match 100.0%; Score 20; DB 16; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 1 IEGR 4

Search completed: April 13, 2003, 03:20:34  
 Job time : 4.22278 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:20:41 ; Search time 0.0997506 Seconds  
(without alignments)  
1179.859 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	1	US-09-377-465A-4
2	20	100.0	4	1	US-07-816-679A-6
3	20	100.0	4	1	US-08-030-731A-2
4	20	100.0	4	1	US-07-854-596B-71
5	20	100.0	4	1	US-08-160-670A-2
6	20	100.0	4	1	US-08-176-500-19
7	20	100.0	4	1	US-08-243-082-20
8	20	100.0	4	1	US-08-170-095B-16
9	20	100.0	4	1	US-08-240-712-2
10	20	100.0	4	1	US-08-240-712-11
11	20	100.0	4	1	US-08-167-035-15
12	20	100.0	4	1	US-08-471-052A-19
13	20	100.0	4	1	US-08-255-272-20
14	20	100.0	4	1	US-08-294-434-3
15	20	100.0	4	1	US-08-468-674B-89
16	20	100.0	4	1	US-08-313-553-4
17	20	100.0	4	1	US-08-110-300A-20
18	20	100.0	4	1	US-08-457-166-3
19	20	100.0	4	1	US-08-396-866-16
20	20	100.0	4	1	US-08-208-887A-15
21	20	100.0	4	1	US-08-313-899A-2
22	20	100.0	4	1	US-08-443-890-2
23	20	100.0	4	1	US-08-443-890-11
24	20	100.0	4	1	US-08-469-486-38
25	20	100.0	4	1	US-08-189-331-19
26	20	100.0	4	1	US-08-153-799-23
27	20	100.0	4	1	US-08-780-571-89

28 20 100.0 4 1 US-08-621-081A-12 Sequence 12, Appli  
29 20 100.0 4 2 US-08-115-519-2 Sequence 2, Appli  
30 20 100.0 4 2 US-08-754-431A-9 Sequence 9, Appli  
31 20 100.0 4 2 US-08-235-515A-19 Sequence 19, Appli  
32 20 100.0 4 2 US-08-471-939-19 Sequence 19, Appli  
33 20 100.0 4 2 US-08-471-800-19 Sequence 15, Appli  
34 20 100.0 4 2 US-08-539-005-15 Sequence 1, Appli  
35 20 100.0 4 2 US-08-203-662-1 Sequence 2, Appli  
36 20 100.0 4 2 US-08-476-014-2 Sequence 2, Appli  
37 20 100.0 4 2 US-07-733-095B-2 Sequence 38, Appli  
38 20 100.0 4 2 US-08-469-658-38 Sequence 6, Appli  
39 20 100.0 4 2 US-07-982-064-6 Sequence 19, Appli  
40 20 100.0 4 2 US-08-471-068-19 Sequence 20, Appli  
41 20 100.0 4 2 US-08-886-642-20 Sequence 18, Appli  
42 20 100.0 4 2 US-08-595-868C-18 Sequence 1, Appli  
43 20 100.0 4 3 US-08-752-892-1 Sequence 4, Appli  
44 20 100.0 4 3 US-08-767-993-4 Sequence 11, Appli  
45 20 100.0 4 3 US-08-722-126A-11

## ALIGNMENTS

RESULT 1  
US-09-377-465A-4  
; Sequence 4, Application US/09377465A  
; Patent No. H002021  
; GENERAL INFORMATION:  
; APPLICANT: Hoskins, JoAnn  
; APPLICANT: Jaskunas, Stanley R  
; APPLICANT: Zhao, Genshi  
; APPLICANT: Rockey, Pamela K  
; TITLE OF INVENTION: NOVEL PENICILLIN BINDING PROTEIN FROM STREPTOCOCCUS  
; FILE REFERENCE: PNEUMONIAE  
; Patent No. H002021  
; CURRENT APPLICATION NUMBER: US/09377,465A  
; CURRENT FILING DATE: 1999-08-19  
; PRIOR APPLICATION NUMBER: 60/100,887  
; PRIOR FILING DATE: 1998-09-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Thrombin Cleavage  
; OTHER INFORMATION: Site  
US-09-377-465A-4

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 2  
US-07-816-679A-6  
; Sequence 6, Application US/07816679A  
; Patent No. 5298599  
; GENERAL INFORMATION:  
; APPLICANT: Rezaie, Alireza  
; APPLICANT: Esmon, Charles T.  
; APPLICANT: Morrissey, James H.  
; TITLE OF INVENTION: Expression and Purification of  
; RECOMBINANT SOLUBLE TISSUE FACTOR  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kilpatrick & Cody

us-10-023-888-22.ra1

Mon Apr 14 09:20:45 2003

```
STREET: 100 Peachtree Street, Suite 3200
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/816,679A
  FILING DATE: 19920103
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/730040
    FILING DATE: 12-JUL-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/292447
    FILING DATE: 30-DEC-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/683682
    FILING DATE: 10-APR-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Pabst, Patrea L.
    REGISTRATION NUMBER: 31,284
    REFERENCE/DOCKET NUMBER: OMR130
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 404-572-6508
      TELEFAX: 404-572-6555
    INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 4 amino acids
        TYPE: AMINO ACID
        STRANDEDNESS: single
        TOPOLOGY: linear
      MOLECULE TYPE: peptide
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FEATURE:
        NAME/KEY: Cleavage-site
        LOCATION: 1..4
        OTHER INFORMATION: /note="Factor xa Cleavage Site"
US-07-816-679A-6

Query Match          100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 3
US-08-030-731A-2
; Sequence 2, Application US/08030731A
; Patent No. 5426036
; GENERAL INFORMATION:
; APPLICANT: Koller, Klaus-Peter
; APPLICANT: Riess, Guenther Johannes
; APPLICANT: Uhlmann, Eugen
; APPLICANT: Wallmeier, Holger
; TITLE OF INVENTION: Processes for the Preparation of Foreign
; TITLE OF INVENTION: Proteins in Streptomyces
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA

STREET: 100 Peachtree Street, Suite 3200
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/07/816,679A
  FILING DATE: 19920103
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/730040
    FILING DATE: 12-JUL-1991
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/292447
    FILING DATE: 30-DEC-1988
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/683682
    FILING DATE: 10-APR-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Pabst, Patrea L.
    REGISTRATION NUMBER: 31,284
    REFERENCE/DOCKET NUMBER: OMR130
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 404-572-6508
      TELEFAX: 404-572-6555
    INFORMATION FOR SEQ ID NO: 6:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 4 amino acids
        TYPE: AMINO ACID
        STRANDEDNESS: single
        TOPOLOGY: linear
      MOLECULE TYPE: peptide
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FEATURE:
        NAME/KEY: Cleavage-site
        LOCATION: 1..4
        OTHER INFORMATION: /note="Factor xa Cleavage Site"
US-07-816-679A-6

Query Match          100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 1.9e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 4
US-08-030-731A-2
; Sequence 71, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplewski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
```

; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/854,596B  
; APPLICATION NUMBER: US/07/854,596B  
; FILING DATE: 03-JUN-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McDonnell, John J  
; REGISTRATION NUMBER: 26,949  
; REFERENCE/DOCKET NUMBER: 92,337  
; TELEPHONE: 312-715-1000  
; TELEFAX: 312-715-1234  
; TELEX: 910-221-5317  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..4  
; OTHER INFORMATION: /note= "Factor Xa cleavable"  
; OTHER INFORMATION: linker"  
; US-07-854-596B-71

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 5  
; Sequence 2, Application US/08160670A  
; Patent No. 5449758  
; GENERAL INFORMATION:  
; APPLICANT: Hartley, James L  
; TITLE OF INVENTION: Protein Size Marker Ladder  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/160,670A  
; FILING DATE: 12/2/93  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Esmond, Robert W.  
; REGISTRATION NUMBER: 32,893  
; REFERENCE/DOCKET NUMBER: 0942.2580000  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: both  
; US-08-160-670A-2  
  
Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 6  
; US-08-176-500-19  
; Sequence 19, Application US/08176500  
; Patent No. 5498538  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/176,500  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/013,416  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-143  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-176-500-19

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 7  
; US-08-243-082-20  
; Sequence 20, Application US/08243082  
; Patent No. 5506120

GENERAL INFORMATION:  
APPLICANT: YAMAMOTO, Hiroaki  
TITLE OF INVENTION: METHOD OF PRODUCING PEPTIDES OR  
PROTEINS  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Spencer, Frank & Schneider  
STREET: 1111 Nineteenth Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/243,082  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/853,754  
FILING DATE: 05-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Schneller, John W.  
REGISTRATION NUMBER: 26,031  
REFERENCE/DOCKET NUMBER: KUWAT 0010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 828-8000  
TELEFAX: (202) 828-8038  
TELEX: SPENCER 64267  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-243-082-20

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
Db 1 IEGR 4

RESULT 8  
US-08-170-095B-16  
Sequence 16, Application US/08170095B  
Patent No. 5563254  
GENERAL INFORMATION:  
APPLICANT: Hoffman, Stephen J.  
APPLICANT: Nagai, Kiyoshi  
TITLE OF INVENTION: Blood Substitutes  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Somatogen, Inc.  
STREET: 2545 Central Avenue  
CITY: Boulder  
STATE: Colorado  
ZIP: 80301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: System 7.0.1  
SOFTWARE: Microsoft Word 5.0a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/170,095B  
FILING DATE: December 20, 1993

CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5563254ak, Henry P.  
REGISTRATION NUMBER: 33200  
REFERENCE/DOCKET NUMBER: Hoffman 2A/CONT2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303-541-3322  
TELEFAX: 303-444-3013  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown to applicant  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
US-08-170-095B-16

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
Db 1 IEGR 4

RESULT 9  
US-08-240-712-2  
Sequence 2, Application US/08240712  
Patent No. 5599907  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, DAVID C.  
APPLICANT: MATHEWS, ANTONY JAMES  
APPLICANT: STETLER, GARY L.  
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
HEMOGLOBINS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brody and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,712  
FILING DATE: 09-MAY-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-240-712-2

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 10  
US-08-240-712-11  
; Sequence 11, Application US/08240712  
; Patent No. 5599907  
; GENERAL INFORMATION:  
; APPLICANT: ANDERSON, DAVID C.  
; APPLICANT: MATHEWS, ANTONY JAMES  
; APPLICANT: STETLER, GARY L.  
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
; TITLE OF INVENTION: HEMOGLOBINS  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Broadway and Neimark  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/240,712  
; FILING DATE: 09-MAY-1994  
; CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/09752  
; FILING DATE: 13-MAY-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: COOPER, IVER P  
; REGISTRATION NUMBER: 28,005  
; REFERENCE/DOCKET NUMBER: ANDERSON-6  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633

INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-240-712-11

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 11  
US-08-167-035-15  
; Sequence 15, Application US/08167035  
; Patent No. 5618691  
; GENERAL INFORMATION:  
; APPLICANT: Schlensing, Joseph  
; APPLICANT: Skolnick, Edward Y.

APPLICANT: Margolis, Benjamin L.  
; TITLE OF INVENTION: NOVEL EXPRESSION CLONING METHOD FOR  
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARYOTIC TYROSINE  
; TITLE OF INVENTION: KINASES AND NOVEL TARGET PROTEINS  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: 10036-2711  
; ZIP: 10036-2711  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/167,035  
; FILING DATE: 16-DEC-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7683-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-167-035-15

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

RESULT 12  
US-08-471-052A-19  
; Sequence 19, Application US/08471052A  
; Patent No. 5625033  
; GENERAL INFORMATION:  
; APPLICANT: Ray, B. K.  
; APPLICANT: Fowlkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 166  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,052A  
; FILING DATE: 06-JUNE-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-179  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-471-052A-19

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
Db 1 IEGR 4

RESULT 13  
US-08-235-272-20  
; Sequence 20, Application US/08255272  
; Patent No. 5627268  
; GENERAL INFORMATION:  
; APPLICANT: Kumar, Ramesh  
; APPLICANT: Sharma, Ajay  
; APPLICANT: Khoury-Christianson, Anastasia  
; APPLICANT: M.  
; TITLE OF INVENTION: Production of Therapeutic Peptides in  
; TITLE OF INVENTION: Transgenic Animals as a Fusion with Hemoglobin  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; APPLICATION NUMBER: US/08/255,272  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; REFERENCE/DOCKET NUMBER: 6794-032  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-235-272-20

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
Db 1 IEGR 4

RESULT 14  
US-08-294-434-3  
; Sequence 3, Application US/08294434  
; Patent No. 5635371  
; GENERAL INFORMATION:  
; APPLICANT: Stout, Jay  
; APPLICANT: Wagner, Fred W.  
; APPLICANT: Coolidge, Thomas R.  
; APPLICANT: Holmquist, Barton  
; TITLE OF INVENTION: CHEMICAL METHOD FOR SELECTIVE  
; TITLE OF INVENTION: MODIFICATION OF THE N- AND/OR C-TERMINAL AMINO ACID  
; TITLE OF INVENTION: ALPHA-CARBON REACTIVE GROUP OF A RECOMBINANT  
; TITLE OF INVENTION: POLYPEPTIDE OR A PORTION THEREOF  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 No. 5635371 West Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,434  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/912,798  
; FILING DATE: 13-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Nelson, Albin J.  
; REGISTRATION NUMBER: 28,650  
; REFERENCE/DOCKET NUMBER: 8648.29-US01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-332-5300  
; TELEFAX: 612-332-9081  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-294-434-3

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IEGR 4  
Db 1 IEGR 4

RESULT 15  
US-08-468-674B-89  
; Sequence 89, Application US/08468674B  
; Patent No. 5639642  
; GENERAL INFORMATION:  
; APPLICANT: Kjeldsen, Thomas B  
; APPLICANT: Vad, Knud  
; TITLE OF INVENTION: SYNTHETIC LEADERS PEPTIDE SEQUENCES  
; NUMBER OF SEQUENCES: 89  
; CORRESPONDENCE ADDRESS:



ADDRESSEE: No. 5639642o No. 5639642disk of No. 5639642th America, Inc.  
STREET: 405 Lexington Avenue, 64th Floor  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Tape  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,674B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/282,852  
FILING DATE: 29-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER: 4085.220-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-878-9655  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-468-674B-89

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 1.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

Search completed: April 13, 2003, 03:29:22  
Job time : 1.09975 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 13, 2003, 03:22:54 ; Search time 0.113051 Seconds  
(without alignments)  
2163.137 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/FCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	4	9	US-09-858-332-8
2	20	100.0	4	9	US-09-832-355A-64
3	20	100.0	4	10	US-09-040-518-6
4	20	100.0	4	10	US-09-998-831-21
5	20	100.0	5	9	US-09-770-102A-51
6	20	100.0	5	9	US-09-259-658-23
7	20	100.0	5	10	US-09-970-308-5
8	20	100.0	5	10	US-09-904-117-9
9	20	100.0	6	9	US-09-899-235-29
10	20	100.0	7	9	US-10-150-262-11
11	20	100.0	8	12	US-10-066-209-6
12	20	100.0	10	12	US-10-066-209-8
13	20	100.0	8	10	US-09-809-517A-2
14	20	100.0	11	9	US-10-092-908-30
15	20	100.0	11	9	US-10-092-908-46
16	20	100.0	15	9	US-09-880-132-62
17	20	100.0	15	10	US-09-880-149-62
18	20	100.0	16	9	US-10-067-790-40
19	20	100.0	16	9	US-10-067-892-40

20	20	100.0	16	9	US-09-539-382-40	Sequence 40, Appl
21	20	100.0	16	9	US-10-067-893-40	Sequence 40, Appl
22	20	100.0	20	10	US-09-205-658-281	Sequence 281, App
23	20	100.0	21	9	US-10-092-908-47	Sequence 47, Appl
24	20	100.0	22	9	US-09-331-631A-16	Sequence 16, Appl
25	20	100.0	29	10	US-09-864-761-48408	Sequence 48408, A
26	20	100.0	31	9	US-10-092-908-48	Sequence 48, Appl
27	20	100.0	37	10	US-09-205-658-257	Sequence 257, App
28	20	100.0	41	9	US-10-092-908-49	Sequence 49, Appl
29	20	100.0	50	10	US-09-957-607-2	Sequence 2, Appli
30	20	100.0	52	9	US-10-091-572-242	Sequence 242, App
31	20	100.0	64	9	US-09-796-692-1068	Sequence 1068, Ap
32	20	100.0	64	9	US-09-796-692-1072	Sequence 1072, Ap
33	20	100.0	64	9	US-09-796-692-1080	Sequence 1080, Ap
34	20	100.0	64	9	US-09-796-692-1091	Sequence 1091, Ap
35	20	100.0	64	9	US-09-796-692-1565	Sequence 1565, Ap
36	20	100.0	64	9	US-09-796-692-1646	Sequence 1646, Ap
37	20	100.0	64	9	US-09-796-692-1738	Sequence 1738, Ap
38	20	100.0	76	9	US-10-104-755-79	Sequence 79, Appl
39	20	100.0	76	9	US-10-104-755-81	Sequence 81, Appl
40	20	100.0	76	9	US-10-104-755-83	Sequence 83, Appl
41	20	100.0	76	9	US-10-125-540-373	Sequence 373, App
42	20	100.0	76	10	US-09-764-870-373	Sequence 373, App
43	20	100.0	78	10	US-09-864-761-38379	Sequence 38379, A
44	20	100.0	86	9	US-09-884-767A-9	Sequence 9, Appli
45	20	100.0	92	9	US-09-738-626-3898	Sequence 3898, Ap

## ALIGNMENTS

## RESULT 1

US-09-858-332-8  
; Sequence 8, Application US/09858332  
; Patent No. US20020164718A1  
; GENERAL INFORMATION:  
; APPLICANT: Tchaga, Grigory S.  
; APPLICANT: Jakhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; FILE OF INVENTION: Using the Same  
; FILE REFERENCE: CLON056CIP  
; CURRENT APPLICATION NUMBER: US/09/858,332  
; CURRENT FILING DATE: 2002-07-02  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-09-858-332-8

Query Match 100.0%; Score 20; DB 9; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

## RESULT 2

US-09-832-355A-64  
; Sequence 64, Application US/09832355A  
; Publication No. US20030027751A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovessdi, Imre  
; APPLICANT: Kessler, Paul

Mon Apr 14 09:20:45 2003

;; TITLE OF INVENTION: VEGF FUSION PROTEINS  
;; FILE REFERENCE: 205654  
;; CURRENT APPLICATION NUMBER: US/09/832,355A  
;; CURRENT FILING DATE: 2001-04-10  
;; NUMBER OF SEQ ID NOS: 126  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 64  
;; LENGTH: 4  
;; TYPE: PRT  
;; ORGANISM: Artificial sequence  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)..(1)  
;; OTHER INFORMATION: Synthetic  
US-09-832-355A-64

Query Match 100.0%; Score 20; DB 9; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 3  
US-09-040-518-6  
;; Sequence 6, Application US/09040518  
;; Patent No. US20010042255A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Karatzas, Costas N.  
;; TITLE OF INVENTION: PRODUCTION OF BIOFILAMENTS IN TRANSGENIC  
;; FILE REFERENCE: 08632/011001  
;; CURRENT APPLICATION NUMBER: US/09/040,518  
;; CURRENT FILING DATE: 1998-03-17  
;; NUMBER OF SEQ ID NOS: 32  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 6  
;; LENGTH: 4  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Designed to act as a recognition site for an  
;; OTHER INFORMATION: enzyme  
US-09-040-518-6

Query Match 100.0%; Score 20; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 4  
US-09-998-831-21  
;; Sequence 21, Application US/09998831  
;; Patent No. US20020119153A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Phillip E. Thorpe  
;; APPLICANT: Rolf A. Brekken  
;; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY  
;; FILE REFERENCE: 4001.002584  
;; CURRENT APPLICATION NUMBER: US/09/998,831  
;; CURRENT FILING DATE: 2001-11-30  
;; PRIOR APPLICATION NUMBER: 09/561,108  
;; PRIOR FILING DATE: 2000-04-28  
;; NUMBER OF SEQ ID NOS: 44  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 21

;; LENGTH: 4  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
;; OTHER INFORMATION: PEPTIDE  
US-09-998-831-21

Query Match 100.0%; Score 20; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 5  
US-09-770-102A-51  
;; Sequence 51, Application US/09770102A  
;; Publication No. US20020197606A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Cyclacel  
;; TITLE OF INVENTION: Compositions and Methods for Monitoring the Modification of Mo  
;; FILE REFERENCE: 10069/1062  
;; CURRENT APPLICATION NUMBER: US/09/770,102A  
;; CURRENT FILING DATE: 2001-01-25  
;; PRIOR APPLICATION NUMBER: US 60/179283  
;; PRIOR FILING DATE: 2000-01-31  
;; NUMBER OF SEQ ID NOS: 57  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 51  
;; LENGTH: 5  
;; TYPE: PRT  
;; ORGANISM: Unknown  
;; FEATURE:  
;; OTHER INFORMATION: Cleavage site  
;; NAME/KEY: SITE  
;; LOCATION: (1)..(5)  
;; OTHER INFORMATION: Cleavage site, Xaa is any amino acid  
;; NAME/KEY: MISC\_FEATURE  
;; LOCATION: (1)..(5)  
;; OTHER INFORMATION: X at position 5 can be any amino acid  
US-09-770-102A-51

Query Match 100.0%; Score 20; DB 9; Length 5;  
Best Local Similarity 100.0%; Pred. No. 2.2e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 6  
US-09-259-658-23  
;; Sequence 23, Application US/09259658  
;; Publication No. US20030032054A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Colyer  
;; APPLICANT: Craig  
;; APPLICANT: Maschio  
;; APPLICANT: Mezna  
;; TITLE OF INVENTION: Compositions And Methods For Monitoring The  
;; FILE REFERENCE: colyer 4256/79245  
;; CURRENT APPLICATION NUMBER: US/09/259,658  
;; CURRENT FILING DATE: 1999-02-26  
;; NUMBER OF SEQ ID NOS: 59  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 23  
;; LENGTH: 5

Mon Apr 14 09:20:45 2003

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa at position 5 can be any amino acid.
; OTHER INFORMATION: Description of Artificial Sequence:Factor Xa
; OTHER INFORMATION: cleavage site.
US-09-259-658-23

Query Match 100.0%; Score 20; DB 9; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 7
US-09-970-308-5
; Sequence 9, Application US/09970308
; Patent No. US20020045193A1
; GENERAL INFORMATION:
; APPLICANT: BRIZZARD, BILLY L.
; APPLICANT: HERNAN, RON
; TITLE OF INVENTION: PURIFICATION OF RECOMBINANT PROTEINS FUSED TO MULTIPLE
; FILE REFERENCE: SGM 6933.2
; CURRENT APPLICATION NUMBER: US/09/970,308
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 09/415,000
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthesized
; NAME/KEY: ACT_SITE
; LOCATION: (5)
; OTHER INFORMATION: X is an amino acid except Proline or Arginine
US-09-970-308-5

Query Match 100.0%; Score 20; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 8
US-09-904-117-9
; Sequence 9, Application US/09904117
; Patent No. US20020151006A1
; GENERAL INFORMATION:
; APPLICANT: MUIR, TOM W.
; APPLICANT: COLE, PHILIP A
; APPLICANT: FRIEDMAN, JEFFREY M.
; APPLICANT: SONDBI, DOLAN
; APPLICANT: SEVERINOV, KONSTANTINE
; TITLE OF INVENTION: METHODS OF LIGATING EXPRESSED PROTEINS
; FILE REFERENCE: 600-1-214CIPB
; CURRENT APPLICATION NUMBER: US/09/904,117
; CURRENT FILING DATE: 2003-07-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/191,890
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/093,990

```

```

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: motif within
; OTHER INFORMATION: linker region
US-09-904-117-9

Query Match 100.0%; Score 20; DB 10; Length 5;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 9
US-09-899-235-29
; Sequence 29, Application US/09899235
; Patent No. US20020173620A1
; GENERAL INFORMATION:
; APPLICANT: HABERMANN, PAUL
; TITLE OF INVENTION: BIFUNCTIONAL FUSION PROTEINS FORMED FROM HIRUDIN AND
; FILE REFERENCE: TAP
; FILE REFERENCE: 02481.1750-00
; CURRENT APPLICATION NUMBER: US/09/899,235
; CURRENT FILING DATE: 2001-07-06
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 6
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-09-899-235-29

Query Match 100.0%; Score 20; DB 9; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 2 IEGR 5

RESULT 10
US-10-150-262-11
; Sequence 11, Application US/10150262.
; Publication No. US20030049264A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, KEITH ALAN
; APPLICANT: DUGGAN, MICHAEL JOHN
; APPLICANT: SHONE, CLIFFORD CHARLES
; TITLE OF INVENTION: CLOSTRIDIAL TOXIN DERIVATIVES ABLE TO MODIFY
; TITLE OF INVENTION: PERIPHERAL
; TITLE OF INVENTION: SENSORY APPARENT FUNCTIONS
; FILE REFERENCE: 023223/0104
; CURRENT APPLICATION NUMBER: US/10/150,262
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: US/09/447,356
; PRIOR FILING DATE: 1999-11-22
; PRIOR APPLICATION NUMBER: 08/945,037
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: GB 9508204.6
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 11

```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Organism: Spacer molecule
US-10-150-262-11
```

```
Query Match          100.0%; Score 20; DB 9; Length 7;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 4 IEGR 7
```

```
RESULT 11
US-10-066-209-6
; Sequence 6, Application US/10066209
; Patent No. US20020115110A1
; GENERAL INFORMATION:
; APPLICANT: Brigham-Burke, Michael R.
; APPLICANT: Young, Peter R.
; TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
; FILE REFERENCE: ANTAGONISTS FOR TUMOR NECROSIS RELATED RECEPTORS TR1 AND TR2
; CURRENT APPLICATION NUMBER: US/10/066,209
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 09/072,993
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/055,513
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 60/056,980
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/057,550
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-066-209-6
```

```
Query Match          100.0%; Score 20; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 3 IEGR 6
```

```
RESULT 12
US-10-066-209-8
; Sequence 8, Application US/10066209
; Patent No. US20020115110A1
; GENERAL INFORMATION:
; APPLICANT: Brigham-Burke, Michael R.
; APPLICANT: Young, Peter R.
; TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
; FILE REFERENCE: ANTAGONISTS FOR TUMOR NECROSIS RELATED RECEPTORS TR1 AND TR2
; CURRENT APPLICATION NUMBER: US/10/066,209
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 09/072,993
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/055,513
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 60/056,980
; PRIOR FILING DATE: 1997-08-26
```

```
; PRIOR APPLICATION NUMBER: 60/057,550
; PRIOR FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-10-066-209-8
```

```
Query Match          100.0%; Score 20; DB 12; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 3 IEGR 6
```

```
RESULT 13
US-09-809-517A-2
; Sequence 2, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/pr
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-2
```

```
Query Match          100.0%; Score 20; DB 10; Length 10;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IEGR 4
    ||||
Db 1 IEGR 4
```

```
RESULT 14
US-10-092-908-30
; Sequence 30, Application US/10092908
; Publication No. US20030040015A1
; GENERAL INFORMATION:
; APPLICANT: Kim, Kwang-Soo
; APPLICANT: Robertson, David
; TITLE OF INVENTION: Methods and Reagents for Identifying
; TITLE OF INVENTION: Compounds and Mutations That Modulate Dopamine
; FILE REFERENCE: Beta-Hydroxylase Activity
; FILE REFERENCE: 04843/097002
; CURRENT APPLICATION NUMBER: US/10/092,908
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,095
; PRIOR FILING DATE: 2001-03-07
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 11
```

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-092-908-30

Query Match 100.0%; Score 20; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 15

US-10-092-908-46  
; Sequence 45, Application US/10092908  
; Publication No. US20030040015A1  
; GENERAL INFORMATION:  
; APPLICANT: Kim, Kwang-Soo  
; APPLICANT: Robertson, David  
; TITLE OF INVENTION: Methods and Reagents for Identifying  
; TITLE OF INVENTION: Compounds and Mutations That Modulate Dopamine  
; TITLE OF INVENTION: Beta-Hydroxylase Activity  
; FILE REFERENCE: 04843/097002  
; CURRENT APPLICATION NUMBER: US/10/092,908  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 60/274,095  
; PRIOR FILING DATE: 2001-03-07  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 46  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 6  
; OTHER INFORMATION: Xaa at position 6 can be Aspartic Acid, Glutamic  
; OTHER INFORMATION: Acid or can be absent.  
US-10-092-908-46

Query Match 100.0%; Score 20; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

Search completed: April 13, 2003, 03:31:01  
Job time : 1.11305 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:19:41 ; Search time 0.844555 Seconds  
(without alignments)  
3053.598 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues  
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main.\*  
1: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pcp.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pcp.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pcp.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pcp.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pcp.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pcp.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pcp.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pcp.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pcp.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pcp.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pcp.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pcp.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pcp.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pcp.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pcp.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pcp.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pcp.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pcp.\*  
24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pcp.\*  
25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pcp.\*  
26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pcp.\*  
27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	4	1	PCT-US02-18014-5
2	20	100.0	4	1	PCT-US02-23375-7
3	20	100.0	4	1	PCT-US02-24119-19
4	20	100.0	4	1	PCT-US02-30093-57
5	20	100.0	4	1	PCT-US94-02388-4
6	20	100.0	4	1	PCT-US96-16032-23

7	20	100.0	4	1	PCT-US97-01652-18	Sequence 18, Appl
8	20	100.0	4	1	PCT-US98-11189-13	Sequence 13, Appl
9	20	100.0	4	1	PCT-US99-01188-26	Sequence 26, Appl
10	20	100.0	4	1	PCT-US99-09258A-13	Sequence 13, Appl
11	20	100.0	4	1	PCT-US99-22756-8	Sequence 8, Appl
12	20	100.0	4	3	US-07-736-847-7	Sequence 7, Appl
13	20	100.0	4	3	US-07-920-449-2	Sequence 2, Appl
14	20	100.0	4	3	US-07-944-328-2	Sequence 2, Appl
15	20	100.0	4	3	US-07-944-328-4	Sequence 4, Appl
16	20	100.0	4	4	US-08-014-096-1	Sequence 1, Appl
17	20	100.0	4	4	US-08-040-743B-6	Sequence 6, Appl
18	20	100.0	4	4	US-08-048-977-1	Sequence 1, Appl
19	20	100.0	4	4	US-08-078-153-15	Sequence 15, Appl
20	20	100.0	4	4	US-08-082-943-1	Sequence 1, Appl
21	20	100.0	4	6	US-08-235-515-19	Sequence 19, Appl
22	20	100.0	4	6	US-08-252-820A-15	Sequence 15, Appl
23	20	100.0	4	6	US-08-252-820B-15	Sequence 15, Appl
24	20	100.0	4	6	US-08-278-774-2	Sequence 2, Appl
25	20	100.0	4	7	US-08-302-892-1	Sequence 1, Appl
26	20	100.0	4	7	US-08-347-893-8	Sequence 8, Appl
27	20	100.0	4	7	US-08-350-528-16	Sequence 16, Appl
28	20	100.0	4	7	US-08-350-530-4	Sequence 4, Appl
29	20	100.0	4	7	US-08-350-530A-4	Sequence 4, Appl
30	20	100.0	4	8	US-08-438-852-2	Sequence 2, Appl
31	20	100.0	4	8	US-08-445-427-6	Sequence 6, Appl
32	20	100.0	4	8	US-08-446-650-3	Sequence 3, Appl
33	20	100.0	4	8	US-08-453-485E-14	Sequence 14, Appl
34	20	100.0	4	8	US-08-468-131-8	Sequence 8, Appl
35	20	100.0	4	8	US-08-471-052-19	Sequence 19, Appl
36	20	100.0	4	8	US-08-471-927-19	Sequence 19, Appl
37	20	100.0	4	8	US-08-477-512A-32	Sequence 32, Appl
38	20	100.0	4	8	US-08-486-817A-32	Sequence 32, Appl
39	20	100.0	4	9	US-08-540-397-38	Sequence 38, Appl
40	20	100.0	4	9	US-08-595-868A-18	Sequence 18, Appl
41	20	100.0	4	10	US-08-626-288-3	Sequence 3, Appl
42	20	100.0	4	10	US-08-643-731-88	Sequence 88, Appl
43	20	100.0	4	10	US-08-680-004-8	Sequence 8, Appl
44	20	100.0	4	12	US-08-866-398-22	Sequence 22, Appl
45	20	100.0	4	13	US-08-909-601-33	Sequence 33, Appl

## ALIGNMENTS

RESULT 1  
PCT-US02-18014-5  
; Sequence 5, Application PC/TUS0218014  
; GENERAL INFORMATION:  
; APPLICANT: University of Utah Research Foundation  
; APPLICANT: Rechsteiner, Martin  
; APPLICANT: Pratt, Gregory  
; APPLICANT: Li, Jun  
; TITLE OF INVENTION: METHODS FOR IDENTIFYING COMPOUNDS WITH  
; TITLE OF INVENTION: THERAPEUTIC POTENTIAL FOR TREATMENT OF CENTRAL  
; TITLE OF INVENTION: NEURODEGENERATIVE DISEASES RESULTING FROM ABNORMAL PROTEIN  
; FILE REFERENCE: 21101.0013p1  
; CURRENT APPLICATION NUMBER: PCT/US02/18014  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: 60/297,332  
; PRIOR FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 4  
; TYPE: PPT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence./note =  
PCT-US02-18014-5

Query Match 100.0% Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 2

PCT-US02-23375-7

; Sequence 7, Application PC/TUS0223375

; GENERAL INFORMATION:

; APPLICANT: Hagen, Frederick S.

; APPLICANT: Woodbury, Richard G.

; APPLICANT: Oort, Pieter J.

; APPLICANT: IcoGen Corporation

; TITLE OF INVENTION: METHODS FOR THE IDENTIFICATION OF PEPTIDYL COMPOUNDS

; TITLE OF INVENTION: INTERACTING WITH EXTRACELLULAR TARGET MOLECULES

; FILE REFERENCE: 17881-7-1PC

; CURRENT APPLICATION NUMBER: PCT/US02/23375

; CURRENT FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 60/306,924

; PRIOR FILING DATE: 2001-07-19

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 7

; LENGTH: 4

; TYPE: PRT

; ORGANISM: Factor Xa recognition sequence

PCT-US02-23375-7

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 3

PCT-US02-24119-19

; Sequence 19, Application PC/TUS0224119

; GENERAL INFORMATION:

; APPLICANT: Haugwitz, Michael

; TITLE OF INVENTION: Method of Detecting Protease Activity in

; FILE REFERENCE: a Cell

; CURRENT APPLICATION NUMBER: PCT/US02/24119

; CURRENT FILING DATE: 2002-07-30

; PRIOR APPLICATION NUMBER: 60/309,312

; PRIOR FILING DATE: 2001-07-31

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 4

; TYPE: PRT

; ORGANISM: human

PCT-US02-24119-19

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 4

PCT-US02-30093-57

; Sequence 57, Application PC/TUS0230093

; GENERAL INFORMATION:

; APPLICANT: University of Rochester

; APPLICANT: Puzas, J. Edward

; TITLE OF INVENTION: COMPOSITIONS AND METHODS INVOLVED IN

; TITLE OF INVENTION: BONE GROWTH

; FILE REFERENCE: 21108.0004P1

; CURRENT APPLICATION NUMBER: PCT/US02/30093

; CURRENT FILING DATE: 2002-09-20

; PRIOR APPLICATION NUMBER: 60/323,987

; PRIOR FILING DATE: 2001-09-20

; NUMBER OF SEQ ID NOS: 78

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 57

; LENGTH: 4

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:/note =

; OTHER INFORMATION: Synthetic Construct

PCT-US02-30093-57

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
| | | |

Db 1 IEGR 4

RESULT 5

PCT-US94-02388-4

; Sequence 4, Application PC/TUS9402388

; GENERAL INFORMATION:

; APPLICANT: TURNER, George J.

; APPLICANT: BETLACH, Mary C.

; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS POLYPEPTIDES

; TITLE OF INVENTION: IN HALOBACTERIA

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Robert Berliner

; STREET: 201 North Figueroa Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90012

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/02388

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Berliner, Robert

; REGISTRATION NUMBER: 20,121

; REFERENCE/DOCKET NUMBER: 5555-206-PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 977-1001

; TELEFAX: (213) 977-1003

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

PCT-US94-02388-4

Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.2e+06; Indels 0; Gaps 0; Mismatches 0;

Matches 4; Conservative 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

## RESULT 6

PCT-US96-16032-23  
; Sequence 23, Application PC/TUS9616032  
; GENERAL INFORMATION:  
; APPLICANT: PENCE  
; TITLE OF INVENTION: Coiled-Coil Heterodimer Methods and  
; TITLE OF INVENTION: Compositions for the Detection and Purification  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/16032  
; FILING DATE: 04-OCT-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/540,397  
; FILING DATE: 06-OCT-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 8900-0109.41  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Factor Xa cleavage site  
PCT-US96-16032-23

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

## RESULT 7

PCT-US97-01652-18  
; Sequence 18, Application PC/TUS9701652  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merchant & Gould  
; STREET: 3100 Norwest Center, 90 S. 7th Street  
; CITY: Minneapolis  
; STATE: Minnesota  
; COUNTRY: United States of America

; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/01652  
; FILING DATE: 04-FEB-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/595,868  
; FILING DATE: 06-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CARTER, Charles G. / BRUESS, Steven C.  
; REGISTRATION NUMBER: 35,093 / 34,130  
; REFERENCE/DOCKET NUMBER: 8648.59W001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612/371-5278 / 612/336-4711  
; TELEFAX: 612/336-4751  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE:  
PCT-US97-01652-18

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

## RESULT 8

PCT-US98-11189-13  
; Sequence 13, Application PC/TUS9811189A  
; GENERAL INFORMATION:  
; APPLICANT: Marcel E. Nimni  
; APPLICANT: Frederick L. Hall  
; APPLICANT: Lingtao Wu  
; APPLICANT: Bo Han  
; APPLICANT: Edwin Shors  
; TITLE OF INVENTION: BONE MORPHOGENETIC PROTEINS AND THEIR  
; TITLE OF INVENTION: USE IN BONE GROWTH  
; FILE REFERENCE: 17972-11  
; CURRENT APPLICATION NUMBER: PCT/US98/11189A  
; CURRENT FILING DATE: 1998-06-02  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Human  
PCT-US98-11189-13

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
||||  
Db 1 IEGR 4

```

RESULT 9
PCT-US99-01188-26
; Sequence 26, Application PC/TUS9901188
; GENERAL INFORMATION:
; APPLICANT: Wittup et al.
; TITLE OF INVENTION: Yeast Cell Surface Display of Proteins and Uses
; FILE REFERENCE: D6061PCT
; CURRENT APPLICATION NUMBER: PCT/US99/01188
; CURRENT FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: 09/009,388
; EARLIER FILING DATE: 1998-01-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: WORD 6.0.1 for Macintosh
; SEQ ID NO 26
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Epitope tag
PCT-US99-01188-26

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 10
PCT-US99-09256A-13
; Sequence 13, Application PC/TUS9909256A
; GENERAL INFORMATION:
; APPLICANT: Maty Aval-Hershkovitz et al.
; TITLE OF INVENTION: GENETICALLY MODIFIED CELLS AND METHODS
; TITLE OF INVENTION: FOR EXPRESSING RECOMBINANT HEPARANASE
; TITLE OF INVENTION: AND METHODS OF PURIFYING SAME
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead* Slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2,
; OPERATING SYSTEM: Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US99/09256A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,618
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 09/071,739
; FILING DATE: May 1, 1998
; APPLICATION NUMBER: 08/922,180
; FILING DATE: September 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 910/16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554

```

```

; TELEX:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US99-09256A-13

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 11
PCT-US99-22756-8
; Sequence 8, Application PC/TUS9922756
; GENERAL INFORMATION:
; APPLICANT: Fronticelli, Clara
; TITLE OF INVENTION: POLYMERIC HEMOGLOBIN MUTANTS
; FILE REFERENCE: 6056-279 PC
; CURRENT APPLICATION NUMBER: PCT/US99/22756
; CURRENT FILING DATE: 1999-09-30
; EARLIER APPLICATION NUMBER: 60/102,640
; EARLIER FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 4
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Factor Xa
; OTHER INFORMATION: recognition sequence
PCT-US99-22756-8

Query Match      100.0%; Score 20; DB 1; Length 4;
Best Local Similarity 100.0%; Pred. No. 4.2e+06;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4
Db 1 IEGR 4

RESULT 12
US-07-736-847-7
; Sequence 7, Application US/07736847
; GENERAL INFORMATION:
; APPLICANT: Blonar, Michael A.
; APPLICANT: Kaelin, William G.
; TITLE OF INVENTION: Plasmids for the Rapid Preparation of
; TITLE OF INVENTION: Modified Proteins
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/736,847
; FILING DATE: 19910729

```

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: DFCI-208  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 861-6240  
TELEFAX: (617) 861-9540  
TELEX: 951794  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-736-847-7

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 13  
US-07-920-449-2  
SEQUENCE 2, Application US/07920449  
GENERAL INFORMATION:  
APPLICANT: LAU, Peter  
APPLICANT: RIOUX, Clement  
TITLE OF INVENTION: LIPOPROTEIN SIGNAL PEPTIDE FUSED TO  
TITLE OF INVENTION: ANTIGENIC POLYPEPTIDES  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/920,449  
FILING DATE: 19920829  
CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: PCT/CA90/00460  
FILING DATE: 27-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 32931/106 SMAB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

US-07-920-449-2

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

RESULT 14  
US-07-944-328-2  
SEQUENCE 2, Application US/07944328  
GENERAL INFORMATION:  
APPLICANT: Sakamoto, Takeshi  
APPLICANT: Takamoto, Kazunori  
APPLICANT: Senda, Toshiya  
APPLICANT: Harada, Yoshinori  
TITLE OF INVENTION: FACTOR Xa LINKER DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Antonelli, Terry, Stout & Kraus  
STREET: 1919 Pennsylvania, Ave., N.W.; Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/944,328  
FILING DATE: 19920914  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-A-3-234430  
FILING DATE: September 13, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Terry, David T.  
REGISTRATION NUMBER: 20,178  
REFERENCE/DOCKET NUMBER: 500.31492XOO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-828-0300  
TELEFAX: 202-828-0380  
TELEX: 440280

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-07-944-328-2

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

RESULT 15  
US-07-944-328-4  
SEQUENCE 4, Application US/07944328  
GENERAL INFORMATION:  
APPLICANT: Sakamoto, Takeshi  
APPLICANT: Takamoto, Kazunori  
APPLICANT: Senda, Toshiya  
APPLICANT: Harada, Yoshinori  
TITLE OF INVENTION: FACTOR Xa LINKER DNA  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Antonelli, Terry, Stout & Kraus  
STREET: 1919 Pennsylvania, Ave., N.W.; Suite 600  
CITY: Washington  
STATE: D.C.

; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/944,328  
; FILING DATE: 19920914  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP-A-3-234430  
; FILING DATE: September 13, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Terry, David T.  
; REGISTRATION NUMBER: 20,178  
; REFERENCE/DOCKET NUMBER: 500.31492XOO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-828-0300  
; TELEFAX: 202-828-0380  
; TELEX: 440280  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-944-328-4

Query Match 100.0%; Score 20; DB 3; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.2e+06;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
Db 1 IEGR 4

Search completed: April 13, 2003, 03:28:06  
Job time : 1.84456 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:21:45 ; Search time 0.202826 Seconds  
(without alignments)  
3053.886 Million cell updates/sec

Title: US-10-023-888-22  
Perfect score: 20  
Sequence: 1 IEGR 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 696924 seqs, 154852082 residues

Total number of hits satisfying chosen parameters: 696924

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		Match	Length	DB	ID	Description
		Match	%					
1	20	100.0		4	1	PCT-US02-33563-8	Sequence 8, Appli	
2	20	100.0		4	1	PCT-US03-29264-35	Sequence 35, Appli	
3	20	100.0		4	1	PCT-US03-04213-6	Sequence 6, Appli	
4	20	100.0		4	1	PCT-US02-37624A-22	Sequence 22, Appli	
5	20	100.0		4	5	US-09-516-055-17	Sequence 17, Appli	
6	20	100.0		4	5	US-09-787-216A-8	Sequence 8, Appli	
7	20	100.0		4	5	US-09-445-576A-104	Sequence 104, App	
8	20	100.0		4	5	US-09-640-198D-28	Sequence 28, Appli	
9	20	100.0		4	5	US-09-858-332A-6	Sequence 6, Appli	
10	20	100.0		4	5	US-09-858-332B-6	Sequence 6, Appli	
11	20	100.0		4	5	US-09-194-223B-150	Sequence 150, App	
12	20	100.0		4	6	US-10-128-587A-70	Sequence 70, Appli	
13	20	100.0		4	6	US-10-274-638-8	Sequence 8, Appli	
14	20	100.0		4	6	US-10-087-775-15	Sequence 15, Appli	
15	20	100.0		4	6	US-10-328-813-10	Sequence 10, Appli	
16	20	100.0		4	6	US-10-158-742A-18	Sequence 18, Appli	
17	20	100.0		4	6	US-10-311-406-5	Sequence 5, Appli	
18	20	100.0		4	6	US-10-343-977-6	Sequence 6, Appli	
19	20	100.0		4	6	US-10-365-095-6	Sequence 6, Appli	
20	20	100.0		4	6	US-10-369-779-45	Sequence 45, Appli	
21	20	100.0		4	6	US-10-373-561-21	Sequence 21, Appli	
22	20	100.0		4	6	US-10-375-913-16	Sequence 16, Appli	
23	20	100.0		5	4	US-08-981-087B-8	Sequence 8, Appli	
24	20	100.0		5	5	US-09-699-314-1	Sequence 1, Appli	
25	20	100.0		5	6	US-10-128-587A-66	Sequence 66, Appli	
26	20	100.0		5	6	US-10-128-587A-69	Sequence 69, Appli	

27 20 100.0 5 6 US-10-128-578B-64 Sequence 64, Appli  
28 20 100.0 6 5 US-09-445-576A-10 Sequence 10, Appli  
29 20 100.0 7 5 US-09-194-223B-71 Sequence 71, Appli  
30 20 100.0 8 6 US-10-360-053-29 Sequence 29, Appli  
31 20 100.0 9 5 US-09-445-576A-77 Sequence 77, Appli  
32 20 100.0 9 5 US-09-445-576A-86 Sequence 86, Appli  
33 20 100.0 9 6 US-10-363-205-106 Sequence 106, App  
34 20 100.0 10 5 US-09-445-576A-82 Sequence 82, Appli  
35 20 100.0 10 6 US-10-130-973A-18 Sequence 18, Appli  
36 20 100.0 13 1 PCT-US02-09017-11 Sequence 11, Appli  
37 20 100.0 13 1 PCT-US02-09017A-11 Sequence 11, Appli  
38 20 100.0 13 5 US-09-935-061-4 Sequence 4, Appli  
39 20 100.0 13 5 US-09-689-516C-11 Sequence 11, Appli  
40 20 100.0 13 5 US-09-194-223B-73 Sequence 73, Appli  
41 20 100.0 13 5 US-09-194-223B-85 Sequence 85, Appli  
42 20 100.0 14 1 PCT-US02-09017-9 Sequence 9, Appli  
43 20 100.0 14 1 PCT-US02-09017-10 Sequence 10, Appli  
44 20 100.0 14 1 PCT-US02-09017A-9 Sequence 9, Appli  
45 20 100.0 14 1 PCT-US02-09017A-10 Sequence 10, Appli

## ALIGNMENTS

RESULT 1  
PCT-US02-33563-8  
; Sequence 8, Application PC/TUS0233563  
; GENERAL INFORMATION:  
; APPLICANT: ZymoGenetics, Inc.  
; TITLE OF INVENTION: DIMERIZED GROWTH FACTOR AND MATERIALS  
; FILE REFERENCE: 01-30FC  
; CURRENT APPLICATION NUMBER: PCT/US02/33563  
; PRIOR FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 60/346,117  
; PRIOR FILING DATE: 2001-10-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: polypeptide, factor Xa cleavage site  
PCT-US02-33563-8

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|||||  
Db 1 IEGR 4

RESULT 2  
PCT-US02-29264-35  
; Sequence 35, Application PC/TUS0229264  
; GENERAL INFORMATION:  
; APPLICANT: Invitrogen Corporation  
; TITLE OF INVENTION: DNA Polymerases and Mutants Thereof  
; FILE REFERENCE: 0942, 536PC01  
; CURRENT APPLICATION NUMBER: PCT/US02/29264  
; CURRENT FILING DATE: 2002-09-16  
; PRIOR APPLICATION NUMBER: US 60/318,903  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3639  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 35  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Factor Xa cleavage site  
PCT-US02-29264-35

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 3

PCT-US03-04213-6  
; Sequence 6, Application PC/TUS0304213  
; GENERAL INFORMATION:  
; APPLICANT: ZymoGenetics, Inc.  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PREPARING DIMERIC GROWTH FACTORS  
; FILE REFERENCE: 01-33PC  
; CURRENT APPLICATION NUMBER: PCT/US03/04213  
; PRIOR FILING DATE: 2003-02-11  
; PRIOR APPLICATION NUMBER: 60/355,882  
; PRIOR FILING DATE: 2002-02-11  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: factor Xa cleavage site  
PCT-US03-04213-6

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 4

PCT-US02-37624A-22  
; Sequence 22, Application PC/TUS0237624A  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, William  
; TITLE OF INVENTION: SOLUBLE GLCNAC PHOSPHOTRANSFERASE  
; FILE REFERENCE: 203515US77  
; CURRENT APPLICATION NUMBER: PCT/US02/37624A  
; CURRENT FILING DATE: 2003-02-27  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
PCT-US02-37624A-22

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 5

US-09-516-055-17  
; Sequence 17, Application US/09516055

; GENERAL INFORMATION:  
; APPLICANT: Karatzas, Costas A.  
; APPLICANT: Huang, Yue-Jin  
; TITLE OF INVENTION: TRANSGENIC ANIMALS THAT PRODUCE ALTERED  
; TITLE OF INVENTION: WOOL  
; FILE REFERENCE: 06632/016002  
; CURRENT APPLICATION NUMBER: US/09/516,055  
; CURRENT FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/149,354  
; PRIOR FILING DATE: 1999-08-18  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-516-055-17

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 6

US-09-787-216A-8  
; Sequence 8, Application US/09787216A  
; GENERAL INFORMATION:  
; APPLICANT: Fronticelli, Clara  
; TITLE OF INVENTION: POLYMERIC HEMOGLOBIN MUTANTS  
; FILE REFERENCE: 6056-279 PC  
; CURRENT APPLICATION NUMBER: US/09/787,216A  
; CURRENT FILING DATE: 2002-09-20  
; PRIOR APPLICATION NUMBER: PCT/US99/22756  
; PRIOR FILING DATE: 1999-09-30  
; PRIOR APPLICATION NUMBER: 60/102,640  
; PRIOR FILING DATE: 1998-10-01  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Factor Xa  
; OTHER INFORMATION: recognition sequence  
US-09-787-216A-8

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

## RESULT 7

US-09-445-576A-104  
; Sequence 104, Application US/09445576A  
; GENERAL INFORMATION:  
; APPLICANT: Borean Pharma A/S  
; TITLE OF INVENTION: Trimerising module  
; FILE REFERENCE: 62032.000004  
; CURRENT APPLICATION NUMBER: US/09/445,576A  
; CURRENT FILING DATE: 2000-07-17  
; NUMBER OF SEQ ID NOS: 104  
; SOFTWARE: PatentIn version 3.1



; SEQ ID NO 104  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fxa protease site  
US-09-445-576A-104

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 8  
US-09-640-198D-28  
; Sequence 28, Application US/09640198D  
; GENERAL INFORMATION:  
; APPLICANT: Russell, Stephen  
; APPLICANT: Kay Whye, Peng  
; TITLE OF INVENTION: System for Monitoring the Location of  
; TITLE OF INVENTION: Transgenes  
; FILE REFERENCE: 07039-295001  
; CURRENT APPLICATION NUMBER: US/09/640,198D  
; PRIOR FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 60/149,168  
; PRIOR FILING DATE: 1999-08-17  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Factor Xa Cleavage Site  
US-09-640-198D-28

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 9  
US-09-858-332A-6  
; Sequence 6, Application US/09858332A  
; GENERAL INFORMATION:  
; APPLICANT: Tchaga, Grigory S.  
; APPLICANT: Jokhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; TITLE OF INVENTION: Using the Same  
; FILE REFERENCE: CLON-056CIP  
; CURRENT APPLICATION NUMBER: US/09/858,332A  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: enterokinase cleavage site  
US-09-858-332A-6

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 10  
US-09-858-332B-6  
; Sequence 6, Application US/09858332B  
; GENERAL INFORMATION:  
; APPLICANT: Tchaga, Grigory S.  
; APPLICANT: Jokhadze, George  
; TITLE OF INVENTION: Metal Ion Affinity Tags and Methods for  
; TITLE OF INVENTION: Using the Same  
; FILE REFERENCE: CLON-056CIP  
; CURRENT APPLICATION NUMBER: US/09/858,332B  
; CURRENT FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: 09/404,017  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/101,867  
; PRIOR FILING DATE: 1998-09-25  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: enterokinase cleavage site  
US-09-858-332B-6

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|  
|  
|  
|  
Db 1 IEGR 4

RESULT 11  
US-09-194-223B-150  
; Sequence 150, Application US/09194223B  
; GENERAL INFORMATION:  
; APPLICANT: COSSET, FRANCOIS-LOIC  
; APPLICANT: VALSESLA, SANDRINE  
; APPLICANT: RUSSELL, STEPHEN J.  
; TITLE OF INVENTION: VIRAL PARTICLES WHICH ARE MASKED OR UNMASKED WITH  
; TITLE OF INVENTION: RESPECT TO A CELL RECEPTOR  
; FILE REFERENCE: 0508-1017  
; CURRENT APPLICATION NUMBER: US/09/194,223B  
; CURRENT FILING DATE: 1999-03-23  
; PRIOR APPLICATION NUMBER: PCT/FR97/00870  
; PRIOR FILING DATE: 1997-05-16  
; PRIOR APPLICATION NUMBER: FR 96.06234  
; PRIOR FILING DATE: 1996-05-20  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 150  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-194-223B-150

Query Match 100.0%; Score 20; DB 5; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

#### RESULT 12

US-10-128-587A-70  
; Sequence 70, Application US/10128587A  
; GENERAL INFORMATION:  
; APPLICANT: Innogenetics N.V.  
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope  
; FILE REFERENCE: 134 PCT  
; CURRENT APPLICATION NUMBER: US/10/128,587A  
; CURRENT FILING DATE: 2002-04-24  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 70  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: adaptor peptide  
US-10-128-587A-70

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

#### RESULT 13

US-10-274-638-8  
; Sequence 8, Application US/10274638  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Margaret D.  
; APPLICANT: Fox, Brian A.  
; TITLE OF INVENTION: DIMERIZED GROWTH FACTOR AND MATERIALS  
; FILE REFERENCE: 01-30  
; CURRENT APPLICATION NUMBER: US/10/274,638  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: 60/346,117  
; PRIOR FILING DATE: 2001-10-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: polypeptide, factor xa cleavage site  
US-10-274-638-8

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

#### RESULT 14

US-10-087-775-15  
; Sequence 15, Application US/10087775  
; GENERAL INFORMATION:  
; APPLICANT: INABA, Niro

; APPLICANT: HORI, Takeya  
; APPLICANT: ITO, Satoru  
; TITLE OF INVENTION: Method for Producing Proteins  
; FILE REFERENCE: 0760-0303P  
; CURRENT APPLICATION NUMBER: US/10/087,775  
; CURRENT FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: JP 2001-60973  
; PRIOR FILING DATE: 2001-03-05  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Intervening peptide region recognized by Factor Xa.  
US-10-087-775-15

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

#### RESULT 15

US-10-328-813-10  
; Sequence 10, Application US/10328813  
; GENERAL INFORMATION:  
; APPLICANT: Osbourne, William R.A.  
; APPLICANT: Ramesh, Nagarajan  
; TITLE OF INVENTION: Compositions and Methods for Treating Diabetes  
; FILE REFERENCE: P-UW 3264  
; CURRENT APPLICATION NUMBER: US/10/328,813  
; CURRENT FILING DATE: 2002-12-23  
; PRIOR APPLICATION NUMBER: US/09/185,852  
; PRIOR FILING DATE: 1998-11-04  
; PRIOR APPLICATION NUMBER: 60/087,660  
; PRIOR FILING DATE: 1998-06-02  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Consensus  
; OTHER INFORMATION: Sequence  
US-10-328-813-10

Query Match 100.0%; Score 20; DB 6; Length 4;  
Best Local Similarity 100.0%; Pred. No. 5.9e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

Db 1 IEGR 4

Search completed: April 13, 2003, 03:30:10  
Job time : 1.20283 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:03:50 ; Search time 0.129676 Seconds  
(without alignments)  
2965.377 Million cell updates/sec

Title: US-10-023-888-22

Perfect score: 20

Sequence: 1 IEGR 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR.73.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	63	2 S24422	4-oxalocrotonate t
2	20	100.0	66	2 C70202	hypothetical prote
3	20	100.0	67	2 A13396	hypothetical prote
4	20	100.0	71	2 D82686	hypothetical prote
5	20	100.0	73	2 T10269	hydroxymethylgluta
6	20	100.0	73	2 H69460	conserved hypothet
7	20	100.0	74	2 S53481	FUN57 protein - ye
8	20	100.0	75	2 AC2441	hypothetical prote
9	20	100.0	75	2 AD2183	hypothetical prote
10	20	100.0	77	2 B83961	acyl-carrier prote
11	20	100.0	77	2 T49620	hypothetical prote
12	20	100.0	80	2 AE1946	hypothetical prote
13	20	100.0	82	2 T01133	hypothetical prote
14	20	100.0	86	2 D71635	hypothetical prote
15	20	100.0	86	2 C97845	ribosomal protein
16	20	100.0	86	2 F85354	50S ribosomal prot
17	20	100.0	88	2 A84568	small nuclear ribo
18	20	100.0	92	2 B90356	probable small nuc
19	20	100.0	94	2 A83523	DNA-directed RNA p
20	20	100.0	101	2 E72458	hypothetical prote
21	20	100.0	102	2 T10099	hypothetical prote
22	20	100.0	102	2 B97736	probable hydroxyme
23	20	100.0	105	2 H87543	hypothetical prote
24	20	100.0	106	2 G96534	conserved hypothet
25	20	100.0	106	2 C90261	hypothetical prote
26	20	100.0	107	2 B97370	hypothetical prote
27	20	100.0	107	2 AB2588	conserved hypothet
28	20	100.0	109	2 JC2440	biphenyl dioxygena
29	20	100.0	109	2 H71142	hypothetical prote

30 20 100.0 109 2 G70317 conserved hypothet  
31 20 100.0 111 2 G75009 hypothetical prote  
32 20 100.0 111 2 AB1383 hypothetical prote  
33 20 100.0 113 2 C69045 hypothetical prote  
34 20 100.0 114 2 C81829 hypothetical prote  
35 20 100.0 114 2 C64511 hypothetical prote  
36 20 100.0 115 2 H72705 hypothetical prote  
37 20 100.0 116 2 H69023 conserved hypothet  
38 20 100.0 119 2 S42881 platelet basic pro  
39 20 100.0 120 2 D69209 hypothetical prote  
40 20 100.0 126 2 S35338 hypothetical prote  
41 20 100.0 127 2 C87640 hypothetical prote  
42 20 100.0 129 1 E69973 hypothetical prote  
43 20 100.0 129 2 F69502 conserved hypothet  
44 20 100.0 129 2 S75257 hypothetical prote  
45 20 100.0 130 2 B71025 hypothetical prote

## ALIGNMENTS

### RESULT 1

S24422

4-oxalocrotonate tautomerase (EC 5.3.2.-) dmpI [similarity] - Pseudomonas putida  
C;Species: Pseudomonas putida  
C;Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 17-Mar-2000  
C;Accession: S24422  
R;Shingler, V.; Powlowski, J.; Marklund, U.  
J. Bacteriol. 174, 711-724, 1992  
A;Title: Nucleotide sequence and functional analysis of the complete phenol/3,4-dimet  
A;Reference number: S24417; MUID:92121108; PMID:1732207  
A;Accession: S24422

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-63 <SHI>

A;Cross-references: EMBL:X60835; NID:g45681; PIDN:CAA43229.1; PID:g45686  
C;Superfamily: 4-oxalocrotonate tautomerase

C;Keywords: aromatic hydrocarbon catabolism; intramolecular oxidoreductase; isomerase

F;2-63/Product: 4-oxalocrotonate tautomerase #status predicted <MAT>

F;2/Active site: Pro #status predicted

Query Match 100.0%; Score 20; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4

|||||

Db 9 IEGR 12

### RESULT 2

C70202

hypothetical protein BB0820 - Lyme disease spirochete  
C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999

C;Accession: C70202

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vukobratovic, B.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943; PMID:9403685

A;Accession: C70202

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-66 <KLE>

A;Cross-references: GB:AE001180; GB:AE000783; NID:g2688755; PIDN:AAAC67174.1; PID:g2688755; A;Experimental source: strain B31

Query Match

Best Local Similarity 100.0%; Score 20; DB 2; Length 66;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|||||  
Db 63 IEGR 66

## RESULT 3

AI3396  
hypothetical protein BME1159 [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
C:Accession: AI3396  
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AI3396  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-67 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL52340.1; PID:g17983135; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME1159  
A:Map position: I

Query Match 100.0%; Score 20; DB 2; Length 67;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|||||  
Db 37 IEGR 40

## RESULT 4

DB2686  
hypothetical protein XF1400 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: DB2686  
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing, Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: DB2686  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-71 <SIM>  
A:Cross-references: GB:AE003971; GB:AE003849; NID:g9106402; PIDN:AAF84209.1; GSPDB:GN00190  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincan, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Rodrigues, V.; Rosa, A.J. de M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF1400

Query Match 100.0%; Score 20; DB 2; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|||||  
Db 24 IEGR 27

## RESULT 5

TI0269  
hydroxymethylglutaryl-CoA reductase (NADPH2) (EC 1.1.1.34) - sea-island cotton (fragm N:Alternate names: 3-hydroxy-3-methylglutaryl coenzyme A reductase  
C:Species: Gossypium barbadense (sea-island cotton)  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Jun-2002  
C:Accession: TI0269  
R:Joost, O.; Blanchini, G.; Bell, A.A.; Benedict, C.R.; Magill, C.W. Mol. Plant Microbe Interact. 8, 880-885, 1995  
A:Title: Differential induction of 3-hydroxy-3-methylglutaryl CoA reductase in two co A:Reference number: Z17010; MUID:96112742; PMID:8664497  
A:Accession: TI0269  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-73 <JOJO>  
A:Cross-references: EMBL:S82272; NID:g1839560; PIDN:AAB47161.1; PID:g1839561  
C:Genetics:  
A:Note: Intron positions not resolved (incomplete sequence)  
C:Function:  
A:Description: catalyzes the NADP-dependent synthesis of mevalonate from 3-hydroxy-3- A:Pathway: isoprenoid biosynthesis  
C:Superfamily: hydroxymethylglutaryl-CoA reductase (NADPH)  
C:Keywords: coenzyme A; isoprenoid biosynthesis; NADP; oxidoreductase

Query Match 100.0%; Score 20; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|||||  
Db 32 IEGR 35

## RESULT 6

H69460  
conserved hypothetical protein AF1689 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: H69460  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E. Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997  
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: H69460  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-73 <KLE>  
A:Cross-references: GB:AE000987; GB:AE000782; NID:g2689310; PIDN:AAB89570.1; PID:g264 C:Superfamily: conserved hypothetical protein MJ1122

Query Match 100.0%; Score 20; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
|||||  
Db 32 IEGR 35

## RESULT 7

S53481  
FUN57 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YAR029w

C:Species: Saccharomyces cerevisiae  
 C>Date: 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 19-Apr-2002  
 C:Accession: S53481  
 R:Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac  
 submitted to the EMBL Data Library, February 1994  
 A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 5  
 A:Reference number: S53458  
 A:Accession: S53481  
 A:Molecule type: DNA  
 A:Residues: 1-74 <BUS>  
 A:Cross-references: EMBL:L28920; NID:g1616966; PID:g456134; GSPDB:GN00001; MIPS:YAR029W  
 C:Genetics:  
 A:Gene: FUN57; MIPS:YAR029W  
 A:Cross-references: SGD:S0000077  
 A:Map position: 1R

Query Match 100.0%; Score 20; DB 2; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 31 IEGR 34

RESULT 8  
 AC2441  
 hypothetical protein asl5083 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AC2441  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AC2441  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-75 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA06782.1; PID:g17134221; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: asl5083

Query Match 100.0%; Score 20; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 IEGR 4  
 Db 22 IEGR 25

RESULT 9  
 AD2183  
 hypothetical protein asr3019 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp.  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
 C:Accession: AD2183  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AD2183  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-75 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA06782.1; PID:g17132113; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120

C:Genetics:  
 A:Gene: asr3019

Query Match 100.0%; Score 20; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 25 IEGR 28

RESULT 10

B83961  
 acyl-carrier protein acpA [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 03-Jun-2002  
 C:Accession: B83961

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
 Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
 A:Reference number: A83650; MUID:20512592; PMID:11058132

A:Accession: B83961

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-77 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06209.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: acpA

C:Superfamily: acyl carrier protein; acyl carrier protein homology

C:Keywords: carrier protein

Query Match 100.0%; Score 20; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 73 IEGR 76

RESULT 11

T49620

hypothetical protein B5022.20 [imported] - Neurospora crassa

C:Species: Neurospora crassa

C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000

C:Accession: T49620

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, May 2000

A:Reference number: Z25022

A:Accession: T49620

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-77 <SCH>

A:Cross-references: EMBL:AL355932; GSPDB:GN00116; NCSP:B5022.20

A:Experimental source: BAC clone B5022; strain OR74A

C:Genetics:

A:Gene: NCSP:B5022.20

A:Map position: 6

A:Introns: 35/2

Query Match 100.0%; Score 20; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IEGR 4  
 Db 14 IEGR 17

RESULT 12

AE1946



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:46:49 ; Search time 494 Seconds  
(without alignments)  
16411.324 Million cell updates/sec

Title: US-10-023-888-1

Perfect score: 3600

Sequence: 1 atggagacagacacactcct.....atcgagacaaattgaagtag 3600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1103.2	30.6	1962	22	ABAA44229
C 2	1103.2	30.6	1962	22	AAK02966
C 3	1103.2	30.6	1962	22	AAI12977
C 4	1103.2	30.6	1962	22	AAI34336
C 5	1103.2	30.6	1962	22	AAI02895
C 6	1103.2	30.6	1962	24	ABS02920
C 7	1101.8	30.6	1125	22	AA49383
C 8	1101.8	30.6	1125	22	AAK15732
C 9	1101.8	30.6	1125	22	AAI22215

C 10	1101.8	30.6	1125	22	AAI47512	Probe #16198 used
C 11	1101.8	30.6	1125	22	AAI07914	Probe #7905 used t
C 12	1101.8	30.6	1125	24	ABS15475	Human genome-deriv
C 13	501	13.9	526	20	AAV86143	EST clone H545. H
C 14	398.4	11.1	537	21	AA43989	Mouse secreted exp
C 15	339	9.4	572	24	ABQ56917	Human colon cancer
C 16	266	7.4	780	22	AAH07289	Human cDNA clone (
C 17	266	7.4	2076	22	AAH14758	Human cDNA sequenc
C 18	226.2	6.3	16800	22	AAK82065	Human immune/haema
C 19	226.2	6.3	16805	22	AAK82067	Human immune/haema
C 20	226.2	6.3	16805	22	AAK82066	Human immune/haema
C 21	209.6	5.8	235	21	AAI19227	Human immune/haema
C 22	208.8	5.8	459	22	ABA57391	Human secreted pro
C 23	208.8	5.8	459	22	AAK05427	Human foetal liver
C 24	208.8	5.8	459	22	AAK31026	Human brain expres
C 25	208.8	5.8	459	22	AAI36936	Human bone marrow
C 26	208.8	5.8	459	24	ABS05776	Probe #5622 used t
C 27	206	5.7	206	22	ABA70017	Human genome-deriv
C 28	206	5.7	206	22	AAK18226	Human foetal liver
C 29	206	5.7	206	22	AAK44122	Human brain expres
C 30	206	5.7	206	22	AAI50130	Human bone marrow
C 31	206	5.7	206	24	ABS18357	Probe #18816 used
C 32	138	3.8	1971	23	ABL04555	Drosophila melanog
C 33	138	3.8	4089	23	ABL04554	Human secreted pro
C 34	113	3.1	114	21	AAK61176	Human immune/haema
C 35	100.2	2.8	234	22	AAK61176	Streptococcus ther
C 36	91.4	2.5	15347	24	ABA01439	Streptococcus ther
C 37	91.4	2.5	28170	24	ABA01447	Streptococcus ther
C 38	76.4	2.1	20555	24	ABA01443	Streptococcus ther
C 39	74.4	2.1	18274	24	ABA01446	Streptococcus ther
C 40	71.4	2.0	135	21	AAA62812	DNA encoding activ
C 41	71.4	2.0	5064	19	AAV69110	Neisseria meningit
C 42	66.6	1.8	5070	22	AAI12839	DNA sequence of ps
C 43	65.6	1.8	156	24	ABA04575	5' fragment of a r
C 44	65.6	1.8	1260	21	AAA98173	S11-VEGF2 constr
C 45	65.6	1.8	1620	21	AAA98174	S11-scVEGF2 constr

## ALIGNMENTS

## RESULT 1

ABAA44229/c

ID ABAA44229 standard; DNA; 1962 BP.

XX

AC ABAA44229;

XX

DT 01-FEB-2002 (first entry)

XX

DE Human breast cell single exon nucleic acid probe #2924.

XX

KW Human; microarray; single exon probe; gene expression; breast;

KW disease; cancer; ss.

XX

OS Homo sapiens.

XX

PN WO200157271-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00662.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX

PI

WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes -

Claim 1; SEQ ID NO 2924; 327bp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1662 BP; 545 A; 383 C; 350 G; 684 T; 0 other;

Query Match 30.6%; Score 1103.2; DB 22; Length 1962;  
 Best Local Similarity 99.7%; Pred. No. 5.6e-283;  
 Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps

QY 1581 AGATCATTTTCATGAATTTGTATAAAGTGATCCCTTCCCAAACACAGACTCACATATATTAT 1640  
 |||||  
 DB 1364 AGATCATTTTCATGAATTTGTATAAAGTGATCCCTTCCCAAACACAGACTCACATATATTAT 1305  
 |||||  
 QY 1641 TCCAAAGGTGAATGCCCTGCCTTATTTTCAGCTTTTCGAGAAGTAGCCAAAAGAGGAGTTGA 1700  
 |||||  
 DB 1304 TCCAAAGGTGAATGCCCTGCCTTATTTTCAGCTTTTCGAGAAGTAGCCAAAAGAGGAGTTGA 1245  
 |||||  
 QY 1701 AGGTGCTATAGTGACAATCCAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1760  
 |||||  
 DB 1244 AGGTGCTATAGTGACAATCCAATAATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1185  
 |||||  
 QY 1761 CATCCACCTCAATAATGCACAGTGGAAATGAATGCCACCACACATACATTTTAATCTCAGTT 1820  
 |||||  
 DB 1184 CATCCACCTCAATAATGCACAGTGGAAATGAATGCCACCACACATACATTTTAATCTCAGTT 1125  
 |||||  
 QY 1821 TCAAAATACAAACGATGAAGAGTTTCAAAATCGAGATACAGTGAGTGAGACACAGGGA 1880  
 |||||  
 DB 1124 TCAAAATACAAACGATGAAGAGTTTCAAAATCGAGATACAGTGAGTGAGACACAGGGA 1065  
 |||||  
 QY 1881 GGGACCAAACTGAATTTCTACGCCCCAGAGGGTTACGAAATTTAGTTAGTCCCATTAAC 1940  
 |||||  
 DB 1064 GGGACCAAACTGAATTTCTACAGCCCAAGGGTTACGAAATTTAGTTAGTCCCATTAAC 1005  
 |||||  
 QY 1941 ACTTCTCCAGAGCGGAAATCCTTTTTCAGGATATTTCCCAAGAAAACGCTTCCCGAA 2000  
 |||||  
 DB 1004 ACTTCTCCAGAGCGGAAATCCTTTTTCAGGATATTTCCCAAGAAAACGCTTCCCGAA 945  
 |||||  
 QY 2001 GTTTAAGACATGATGTTTAATCAACAGGAGGCCCAGGAAGAGTGGAATTCCTCCCT 2060  
 |||||  
 DB 944 GTTTAAGACATGATGTTTAATCAACAGGAGGCCCAGGAAGAGTGGAATTCCTCCCT 885  
 |||||  
 QY 2061 GGTAAATATTTCACCTCTCCAAAAGCCCGCTTGAGTCTCAATACCTTGGATTGCA 2120  
 |||||  
 DB 884 GGTAAATATTTCACCTCTCCAAAAGCCCGCTTGAGTCTCAATACCTTGGATTGCA 825  
 |||||  
 QY 2121 ACTGGAACATGGAGACATCACCTTGAAGGATACAAATTTGTCCAAAGTAGCCTTGCTGAG 2180  
 |||||  
 DB 824 ACTGGAACATGGAGACATCACCTTGAAGGATACAAATTTGTCCAAAGTAGCCTTGCTGAG 765  
 |||||



xx Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 PS Example 4; SEQ ID NO: 2957; 650pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 XX  
 XX Sequence 1962 BP; 545 A; 383 C; 350 G; 684 T; 0 other;

Query Match 30.6%; Score 1103.2; DB 22; Length 1962;  
 Best Local Similarity 99.7%; Pred. No. 5.6e-283;  
 Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1581 AGATCATTTTCATGAATTCGTATAAAGTGTATCCCTTCTCCCAACACAGACTCAGTATATAT 1640  
 DB 1364 AGATCATTTTCATGAATTCGTATAAAGTGTATCCCTTCTCCCAACACAGACTCAGTATATAT 1305  
 QY 1641 TCCAAAGGTGAATGCTCCCTTATTTTCAGCTTTGCGAAGTAGCCAAAGAGGAGTTGA 1700  
 DB 1304 TCCAAAGGTGAATGCTCCCTTATTTTCAGCTTTGCGAAGTAGCCAAAGAGGAGTTGA 1245  
 QY 1701 AGGTGCTATAGTACATCCCAATATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1760  
 DB 1244 AGGTGCTATAGTACATCCCAATATTCGACATGCTTCTATTGCCAACAGTGGAAAC 1185  
 QY 1761 CATCCACCTCATATGACAGTGGAGTGAATGCCACCAATACATATTTAATCTCAGCTT 1820  
 DB 1184 CATCCACCTCATATGACAGTGGAGTGAATGCCACCAATACATATTTAATCTCAGCTT 1125  
 QY 1821 TCAAAATACAAAGTGAAGTTCAAAATGCAGATACAGTGGAGTGGACAAAGGA 1880  
 DB 1124 TCAAAATACAAAGTGAAGTTCAAAATGCAGATACAGTGGAGTGGACAAAGGA 1065  
 QY 1881 GGGACCAAACTGAATTCACGGCCCAAGGTTTACCAAAATTTAGTTAGTCCCATAC 1940  
 DB 1064 GGGACCAAACTGAATTCACGGCCCAAGGTTTACCAAAATTTAGTTAGTCCCATAC 1005  
 QY 1941 ACTTCTCCAGAGCGGAAATCCTTTTGGAGTATTTCCCAAGAAACGCTTCCCGAA 2000  
 DB 1004 ACTTCTCCAGAGCGGAAATCCTTTTGGAGTATTTCCCAAGAAACGCTTCCCGAA 945  
 QY 2001 GTTTAAGAGACATGATGTTAACTCAACAGAGAGCCGAGGAGTGAATTTCCCT 2060  
 DB 944 GTTTAAGAGACATGATGTTAACTCAACAGAGAGCCGAGGAGTGAATTTCCCT 885  
 QY 2061 GGTAAATATTTCACTCTCCCAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA 2120  
 DB 884 GGTAAATATTTCACTCTCCCAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA 825  
 QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCCAAAGTGCCTTGTGAG 2180  
 DB 824 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCCAAAGTGCCTTGTGAG 765  
 QY 2181 ATCATTTCTGATGAATCAGACATGCTTAAATAAATAAATCAAGCTATATACAGATGA 2240  
 DB 764 ATCATTTCTGATGAATCAGACATGCTTAAATAAATAAATCAAGCTATATACAGATGA 705  
 QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCATAAAGCATCTTGGCAAA 2300  
 DB 704 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCATAAAGCATCTTGGCAAA 645  
 QY 2301 CAGCTTAGGAGTCTGAAAGATTCAGAGTTGACATTTTCCCTGCAAGTGAATTAAGT 2360  
 DB 644 CAGCTTAGGAGTCTGAAAGATTCAGAGTTGACATTTTCCCTGCAAGTGAATTAAGT 595  
 QY 2361 GAATGGTCAATGACCAAGGTCAGAAATCCACCCCTGGACTTGGAGACCAACAGATTTAG 2420

DB 584 GAATGGTCAATGACACAGGTCAGAAATCCACCCCTGGACTTGGAGACCAAGCAATTTAG 525  
 QY 2421 AGTGAAGAACTCACACCCCAAAAACCATAGCGGAAATGTGACAAAAGAAAAAGCCCCATC 2480  
 DB 524 AGTGAAGAACTCACACCCCAAAAACCATAGCGGAAATGTGACAAAAGAAAAAGCCCCATC 465  
 QY 2481 TCTGATTCTTCCACTGGAAAGCCAGATGACAAAAGAAAAAATCAGAGGAAAGAAAA 2540  
 DB 464 TCTGATTCTTCCACTGGAAAGCCAGATGACAAAAGAAAAAATCAGAGGAAAGAAAA 405  
 QY 2541 AGAGAACAGTGAATGGAGGAAAAATGCTGAAAAATCAGATAGCGCTTACTGAAGTGTACT 2600  
 DB 404 AGAGAACAGTGAATGGAGGAAAAATGCTGAAAAATCAGATAGCGCTTACTGAAGTGTACT 345  
 QY 2601 TGGAAAGAACTGCGAGCATTTACAGATAGTACTTGGGCTTTTGGCCATGGGAGAAAA 2660  
 DB 344 TGGAAAGAACTGCGAGCATTTACAGATAGTACTTGGGCTTTTGGCCATGGGAGAAAA 285  
 QY 2661 AAGTATTTCTAGATCTTCTTCGACGAA 2688  
 DB 284 AAGTATTTCTAGATCTTCTTCGACGTA 257

RESULT 3

AA112977/c

ID AA112977 standard; DNA; 1962 BP.

XX AA112977;

XX AA112977;

DT 12-OCT-2001 (first entry)

XX Probe #2910 for gene expression analysis in human cervical cell sample.

XX Probe: human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID No 2910; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
 CC (SENP). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO

1

Db 1244 AGGTGCCTATAGTGAACATCCATATTCGACATGCTCTATTGCCAACAAGTGGAAC 1185  
 QY 1761 CATCCACCTCATATGACAGTGGAGTAAATGCCACCAATACATATTTAATCTCACGTT 1820  
 Db 1184 CATCCACCTCATATGACAGTGGAGTAAATGCCACCAATACATATTTAATCTCACGTT 1125  
 QY 1821 TCAAAATCAAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAAGGA 1880  
 Db 1124 TCAAAATCAAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAAGGA 1055  
 QY 1881 GGGACCAAACTGAATCTACGGCCAGAGGGTTACGAAATTTAGTTAGTCCCATAC 1940  
 Db 1064 GGGACCAAACTGAATCTACGGCCAGAGGGTTACGAAATTTAGTTAGTCCCATAC 1005  
 QY 1941 ACTTCTTCAGAGCGGAAATCTTTTTCAGATATTCCTCCAAAGAAACGCTTCCGAA 2000  
 Db 1004 ACTTCTTCAGAGCGGAAATCTTTTTCAGATATTCCTCCAAAGAAACGCTTCCGAA 945  
 QY 2001 GTTTAAGACATGATGTTAACTCAACAGGAGAGCCAGGAGGTGAAATTTCCCT 2060  
 Db 944 GTTTAAGACATGATGTTAACTCAACAGGAGAGCCAGGAGGTGAAATTTCCCT 885  
 QY 2061 GGTAAATATTCACCTCTCCAAAGAGCGCCAGTTCAGTCTCAATACCTTGGATTGCA 2120  
 Db 884 GGTAAATATTCACCTCTCCAAAGAGCGCCAGTTCAGTCTCAATACCTTGGATTGCA 825  
 QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCACAGTTCAGCTTGCAG 2180  
 Db 824 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCACAGTTCAGCTTGCAG 765  
 QY 2181 ATCATTTCATGATCACTCAAGATGCTAAATTAATAAATCAAGCTATATAACAGATGA 2240  
 Db 764 ATCATTTCATGATCACTCAAGATGCTAAATTAATAAATCAAGCTATATAACAGATGA 705  
 QY 2241 AACAAATCACAGTTCCTGCTCCACAGGAAACAGTTCATATAAGCATCTTGCCAAA 2300  
 Db 704 AACAAATCACAGTTCCTGCTCCACAGGAAACAGTTCATATAAGCATCTTGCCAAA 645  
 QY 2301 CAGCTTAGGAGTGTCTGAAAGATTCGAGAGTTCAGTTCCTTCAGTTCAGTTCAGTTCAGT 2360  
 Db 644 CAGCTTAGGAGTGTCTGAAAGATTCGAGAGTTCAGTTCCTTCAGTTCAGTTCAGTTCAGT 585  
 QY 2361 GAATGGTTCATGACAGGTCAGATCCACCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCAG 2420  
 Db 584 GAATGGTTCATGACAGGTCAGATCCACCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCAG 525  
 QY 2421 AGTGGAACTCACACCCAAAAACCATAGCGGAAATGTGACAAAGAAAGCCCAATC 2480  
 Db 524 AGTGGAACTCACACCCAAAAACCATAGCGGAAATGTGACAAAGAAAGCCCAATC 465  
 QY 2481 TCTGATTTCCACTGGAAGCCAGATGACAAAGAAAGAAATCAACAGGGAAGAA 2540  
 Db 464 TCTGATTTCCACTGGAAGCCAGATGACAAAGAAAGAAATCAACAGGGAAGAA 405  
 QY 2541 AGAGACAGTGAATGGAGGAAATGCTGAAATCACATAGCGTTCAGTTCAGTTCAGTTCAGT 2600  
 Db 404 AGAGACAGTGAATGGAGGAAATGCTGAAATCACATAGCGTTCAGTTCAGTTCAGTTCAGT 345  
 QY 2601 TGGAGAAAGCTGCACATACAGATAGTTACTTGGGCTTTTTCGATGGGAGAA 2660  
 Db 344 TGGAGAAAGCTGCACATACAGATAGTTACTTGGGCTTTTTCGATGGGAGAA 285  
 QY 2661 AAAGTATTTCTAGATCTTCGAGCA 2688  
 Db 284 AAAGTATTTCTAGATCTTCGAGCA 257

RESULT 5  
 AA102895/c  
 ID AA102895 standard; DNA; 1962 BP.  
 XX  
 AC AA102895;

XX DT 09-OCT-2001 (first entry)  
 XX DE Probe #2886 used to measure gene expression in human breast sample.  
 XX KW Probe; human; breast disease; breast cancer; development disorder; ss;  
 XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 OS Homo sapiens.  
 PN WO200157270-A2.  
 XX PD 09-AUG-2001.  
 XX PF 29-JAN-2001; 2001WO-US00661.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WIPI; 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 in a human breast -  
 PS Claim 25; SEQ ID No 2886; 322pp; English.  
 XX The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast,  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 1962 BP; 545 A; 383 C; 350 G; 684 T; 0 other;

Query Match 30.6%; Score 1103.2; DB 22; Length 1962;  
 Best Local Similarity 99.7%; Pred. No. 5.6e-283;  
 Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1581 AGATCATTTTCATGAATTCATATAAGTATCTCTCCCAACACAGCTCCTACTATATAT 1640  
 Db 1364 AGATCATTTTCATGAATTCATATAAGTATCTCTCTCCCAACACAGCTCCTACTATATAT 1305  
 QY 1641 TCCAAAGGTGAATGCCTGCCTTATTTTCAGCTTTTCAGAAAGTACCAAGAGGAGTTGA 1700  
 Db 1304 TCCAAAGGTGAATGCCTGCCTTATTTTCAGCTTTTCAGAAAGTACCAAGAGGAGTTGA 1245  
 QY 1701 AGGTGCTATAGTGAATTCCTCAATTAATTCGACATGCTTCTATTGCCAACAAGTGA 1760  
 Db 1244 AGGTGCTATAGTGAATTCCTCAATTAATTCGACATGCTTCTATTGCCAACAAGTGA 1185  
 QY 1761 CATCCACCTCATATGACAGTGGAGTAAATGCCACCAATACATATTTAATCTCACGTT 1820  
 Db 1184 CATCCACCTCATATGACAGTGGAGTAAATGCCACCAATACATATTTAATCTCACGTT 1125  
 QY 1821 TCAAAATCAAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAAGGA 1880

Db 1124 TCAAAATCAACAGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAAGGA 1065  
 QY 1881 GGGACCAAACTGAATTCACGCCAGAGGTTACGAAATTTAGTTAGTCCATAAC 1940  
 Db 1064 GGGACCAAACTGAATTCACGCCAGAGGTTACGAAATTTAGTTAGTCCATAAC 1005  
 QY 1941 ACTTCTTCAGAGGGGAAATCTTTTGGAGATATCCCAAGAAACACGCTTCCCGAA 2000  
 Db 1004 ACTTCTTCAGAGGGGAAATCTTTTGGAGATATCCCAAGAAACACGCTTCCCGAA 945  
 QY 2001 GTTTAAGAGACATGATTTAACTCAACAGAGAGCCAGAGAGGTGAAATTCGCC 2060  
 Db 944 GTTTAAGAGACATGATTTAACTCAACAGAGAGCCAGAGAGGTGAAATTCGCC 885  
 QY 2061 GGTAAATATTTTCACTCCCTTCCAAAAGAGCCAGTTGAGTCTCAATACCTTGGATTGCA 2120  
 Db 884 GGTAAATATTTTCACTCCCTTCCAAAAGAGCCAGTTGAGTCTCAATACCTTGGATTGCA 825  
 QY 2121 ACTGGAACATGAGAGACATCACTTTGAAAGGATACAAATTTGTCACAGCTTCTGCTGAG 2180  
 Db 824 ACTGGAACATGAGAGACATCACTTTGAAAGGATACAAATTTGTCACAGCTTCTGCTGAG 765  
 QY 2181 ATCAATTTCTGATGAAGTCACTGCTTAAATAAATAAATCAAGCTATAATAACAGATGA 2240  
 Db 764 ATCAATTTCTGATGAAGTCACTGCTTAAATAAATAAATCAAGCTATAATAACAGATGA 705  
 QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGAGAAACAGAGTTTCAATAAGCATCTTCCCAA 2300  
 Db 704 AACAAATGACAGTTTGGTGGCTCCACAGAGAAACAGAGTTTCAATAAGCATCTTCCCAA 645  
 QY 2301 CAGCTTAGAGTGTCTGAAGATGTCAGAGTTGACTTTTCTGTCAGTGAGTAAAGT 2360  
 Db 644 CAGCTTAGAGTGTCTGAAGATGTCAGAGTTGACTTTTCTGTCAGTGAGTAAAGT 585  
 QY 2361 GAATGTCATGACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 2420  
 Db 584 GAATGTCATGACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 525  
 QY 2421 AGTGGAACTCACACCCCAAAACCAATAGCGGAATGTGACAAAAGAAAGCCCCATC 2480  
 Db 524 AGTGGAACTCACACCCCAAAACCAATAGCGGAATGTGACAAAAGAAAGCCCCATC 465  
 QY 2481 TCTGATTTCTCACTGAAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAAA 2540  
 Db 464 TCTGATTTCTCACTGAAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAAA 405  
 QY 2541 AGAAGACATGATGAGGAAATGCTGAAATCAGATAGGCGTTACTGAAGTGTACT 2600  
 Db 404 AGAAGACATGATGAGGAAATGCTGAAATCAGATAGGCGTTACTGAAGTGTACT 345  
 QY 2601 TGGAGAAAGCTGACAGATACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAA 2660  
 Db 344 TGGAGAAAGCTGACAGATACAGATAGTTACTTGGGCTTTTGGCATGGGAGAAAA 285  
 QY 2661 AAAGTATTTCTAGATCTTCTCGACGAA 2688  
 Db 284 AAAGTATTTCTAGATCTTCTCGACGTA 257  
 RESULT 6  
 ID ABS02920 standard; DNA; 1962 BP.  
 XX ABS02920;  
 AC ABS02920;  
 XX 19-AUG-2002 (first entry)  
 DE Human genome-derived single exon probe from lung SEQ ID No 2911.  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 pulmonary histiocytosis; lymphangiomyomatosis; Karagener syndrome;  
 pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 primary ciliary dyskinesia; pulmonary hypertension;  
 hyaline membrane disease.  
 OS Homo sapiens.  
 PN WO200186003-A2.  
 XX 15-NOV-2001.  
 XX 30-JAN-2001; 2001WO-US00665.  
 XX 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 PS Claim 1; SEQ ID No 2911; 634pp; English.  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1962 BP; 545 A; 383 C; 350 G; 684 T; 0 other;  
Query Match 30.6%; Score 1103.2; DB 24; Length 1962;  
Best Local Similarity 99.7%; Pred. No. 5.6e-283;  
Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1581 AGATCAATTTTCATGAATTTGATATAAGTGTATCTTCTCCCAACACAGCTCAGTATATATAT 1640  
DB 1364 AGATCAATTTTCATGAATTTGATATAAGTGTATCTTCTCCCAACACAGCTCAGTATATATAT 1305  
QY 1641 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTTGCAGAAGTAGCCAAAGAGGAGTTGA 1700  
DB 1304 TCCAAAAGGTGAATGCTGCTTATTTTCAGCTTTTGCAGAAGTAGCCAAAGAGGAGTTGA 1245  
QY 1701 AGTGCCCTATAGTGACAATCCCAATATTCGACATGCTTCTATTTGCCAACAGAGTGAAAAAC 1760  
DB 1244 AGTGCCCTATAGTGACAATCCCAATATTCGACATGCTTCTATTTGCCAACAGAGTGAAAAAC 1185  
QY 1761 CATCCACCTCATATATGCACAGTGGATGAATGCCACCAATACATATATTTTAATCTCAGCTT 1820  
DB 1184 CATCCACCTCATATATGCACAGTGGATGAATGCCACCAATACATATTTTAATCTCAGCTT 1125  
QY 1821 TCAAAATACAAACGATGAAGTTCAAAATGCAGATACAGTGGAGGTGGACACAAAGGA 1880  
DB 1124 TCAAAATACAAACGATGAAGTTCAAAATGCAGATACAGTGGAGGTGGACACAAAGGA 1065  
QY 1881 GGGACCAAACTGAATTTCTACGCGCCAGAGGTTTACGAAATTTAGTTAGTCCCATTAAC 1940  
DB 1064 GGGACCAAACTGAATTTCTACGCGCCAGAGGTTTACGAAATTTAGTTAGTCCCATTAAC 1005  
QY 1941 ACTTCTTCAGAGGGGAAATCCTTTTGGAGATATTCCTCCCAAGAAACGCTTCCCGAA 2000  
DB 1004 ACTTCTTCAGAGGGGAAATCCTTTTGGAGATATTCCTCCCAAGAAACGCTTCCCGAA 945  
QY 2001 GTTAAAGACATGATGTTAACTCAACAGGAGAGCCAGGAGGTGAAATTTCCCT 2060  
DB 944 GTTAAAGACATGATGTTAACTCAACAGGAGAGCCAGGAGGTGAAATTTCCCT 885  
QY 2061 GGTAAATATTTTCACTCTTCCAAAGAGCCAGTTCAGTCTCAATACCTTGGATTGCA 2120  
DB 884 GGTAAATATTTTCACTCTTCCAAAGAGCCAGTTCAGTCTCAATACCTTGGATTGCA 825  
QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCAGGTCGCTGAG 2180  
DB 824 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCAGGTCGCTGAG 765  
QY 2181 ATCAATTTCTGATGACTCAGACATGCTTAAATTAATAAATCAAGCTATATACAGATGA 2240  
DB 764 ATCAATTTCTGATGACTCAGACATGCTTAAATTAATAAATCAAGCTATATACAGATGA 705  
QY 2241 AACAAATCACAGTTTGGTGGTCCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAA 2300  
DB 704 AACAAATCACAGTTTGGTGGTCCACAGGAAACAGGTTTCATAAAGCATCTTGCCAAA 645  
QY 2301 CAGCTTAGGAGTGTCTGAAAGATTGACAGAGTTTGACTTTTCCGCAAGTGAAGT 2360  
DB 644 CAGCTTAGGAGTGTCTGAAAGATTGACAGAGTTTGACTTTTCCGCAAGTGAAGT 585  
QY 2361 GAATGGTATGACACAGGTCAGATCAACCCCTGGACTTGGAGACACAGCAAGATTAG 2420  
DB 584 GAATGGTATGACACAGGTCAGATCAACCCCTGGACTTGGAGACACAGCAAGATTAG 525  
QY 2421 AGTGGAACTCACACCAAAACCATAGCCGGAATGTGACAAAGAAAGCCCCCATC 2480  
DB 524 AGTGGAACTCACACCAAAACCATAGCCGGAATGTGACAAAGAAAGCCCCCATC 465  
QY 2481 TCTGATTTCTCACTGGAAGCCAGATGACAAAGAAAGAAATCAAGGGAAGAAA 2540  
DB 464 TCTGATTTCTCACTGGAAGCCAGATGACAAAGAAAGAAATCAAGGGAAGAAA 405  
QY 2541 AGAGACAGTAGAATGGAGAAAATGCTGAAAATCACATAGGCGTTACTGAAGTCTTACT 2600  
|||||

Db 404 AGAGAACAGTAGAATGGAGGAAAATGCTGAAAATCACATAGCGTTACTGAGTCTTACT 345  
QY 2601 TGAAGAAAAGCTCGACATTTACACATAGTTACTTGGCTTTTGGCCATGGGAGAAAA 2660  
Db 344 TGAAGAAAAGCTCGACATTTACACATAGTTACTTGGCTTTTGGCCATGGGAGAAAA 285  
QY 2661 AAAAGTATTTCTAGATCTTCTCGACGAA 2688  
Db 284 AAAAGTATTTCCAAGATCTTCTCGACGTA 257  
|||||

RESULT 7  
ABA49383/c  
ID ABA49383 standard; DNA; 1125 BP.  
XX ABA49383;  
AC  
XX  
AC  
XX  
DT 01-FEB-2002 (first entry)  
XX Human breast cell single exon nucleic acid probe #8078.  
DE Human; microarray; single exon probe; gene expression; breast;  
XX disease; cancer; ss.  
KW  
XX Homo sapiens.  
OS  
XX  
PN WO200157271-A2.  
XX  
XX  
PD 09-AUG-2001.  
XX  
XX  
PF 30-JAN-2001; 2001WO-US00662.  
XX  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-496933/54.  
XX  
XX New spatially-addressable set of single exon nucleic acid probes,  
XX useful for measuring gene expression in sample derived from human  
XX breast, comprises number of single exon nucleic acid probes -  
XX  
XX Claim 4; SEQ ID NO 8078; 327pp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon  
nucleic acid probes for measuring gene expression in a sample derived  
from human breast and BT 474 cells. The method involves contacting  
the probes with a collection of detectably labelled nucleic acids  
derived from mRNA of human breast, and then measuring the label  
bound to each probe of the microarray. The probes are useful for  
verifying the expression of regions of genomic DNA predicted to  
encode proteins. They are useful for gene discovery, and for  
determining predisposition and/or prognosing breast disease. Gene  
expression analysis is useful for assessing the toxicity of chemical  
agents on cells. The microarray of this invention presents a far greater  
diversity of probes for measuring gene expression, with far less bias  
than expressed sequence tag microarrays. The method is suitable for  
rapid production of functional information from genomic sequence. The  
present sequence is a single exon nucleic acid probe of the invention.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

Mon Apr 14 09:20:31 2003

us-10-023-888-1.l.rng

Query Match 30.6%; Score 1101.8; DB 22; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-282;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1581 AGATCAATTTTCATGAATGTTATTAAGTGTATCTCTCCCAACACAGACTCAGTATATATAT 1640  
Db 1105 AGATCAATTTTCATGAATGTTATTAAGTGTATCTCTCCCAACACAGACTCAGTATATAT 1046

QY 1641 TCCAAAGGTGATGCTGCTGCTTATTCAGCTTTCGCAAGTAGCCAAAGAGAGGTGA 1700  
Db 1045 TCCAAAGGTGATGCTGCTGCTTATTCAGCTTTCGCAAGTAGCCAAAGAGAGGTGA 986

QY 1701 AGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTCGCAACAGAGTGA 1760  
Db 985 AGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTCGCAACAGAGTGA 926

QY 1761 CATCCACCTCAATGACAGAGTGAATGATGCCACCAATACATTTTAAATCTCAGCTT 1820  
Db 925 CATCCACCTCAATGACAGTGAATGATGCCACCAATACATTTTAAATCTCAGCTT 866

QY 1821 TCAAAATACAAAGCATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGA 1880  
Db 865 TCAAAATACAAAGCATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGGACACAAAGGA 806

QY 1881 GGGACCAAACTGAATTTCTAGGCCCCAGAGGTTACGAAATTTAGTTAGTCCCATAC 1940  
Db 805 GGGACCAAACTGAATTTCTAGGCCCCAGAGGTTACGAAATTTAGTTAGTCCCATAC 746

QY 1941 ACTCTTCCAGAGCGGAAATCTTTTGGAGTATTTCCCAAGAAAAACGCTTCCCGAA 2000  
Db 745 ACTCTTCCAGAGCGGAAATCTTTTGGAGTATTTCCCAAGAAAAACGCTTCCCGAA 686

QY 2001 GTTTAAGAGCATGATGTTAACTCAACAGAGAGGAGCCAGAGAGTGAATTTCCCT 2060  
Db 685 GTTTAAGAGCATGATGTTAACTCAACAGAGAGGAGCCAGAGAGTGAATTTCCCT 626

QY 2061 GGTAAATATTCATCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA 2120  
Db 625 GGTAAATATTCATCTCTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA 566

QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCAAGTCAGCCTTGTGAG 2180  
Db 565 ACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCAAGTCAGCCTTGTGAG 506

QY 2181 ATCAATTTCTGATGACATCAGAGTCTAAATTAATAAATCAAGCTATAATAACAGATGA 2240  
Db 505 ATCAATTTCTGATGACATCAGAGTCTAAATTAATAAATCAAGCTATAATAACAGATGA 446

QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGAAAAACAGGTTTCATAAAAGCATCTTGGCAAA 2300  
Db 445 AACAAATGACAGTTTGGTGGCTCCACAGAAAAACAGGTTTCATAAAAGCATCTTGGCAAA 386

QY 2301 CAGCTTAGAGTGTCTGAAAGATTGACAGATTGACATTTTCTCAGTGTGAGTGAAGT 2360  
Db 385 CAGCTTAGAGTGTCTGAAAGATTGACAGATTGACATTTTCTCAGTGTGAGTGAAGT 326

QY 2361 GAATGCTATGACAGAGGTGAGATCCACCTTGGACTTGGAGACCCAGCAAGATTAG 2420  
Db 325 GAATGCTATGACAGAGGTGAGATCCACCTTGGACTTGGAGACCCAGCAAGATTAG 266

QY 2421 AGTGAAACTCACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATC 2480  
Db 265 AGTGAAACTCACACCCCAAAACCATAGGCGGAAATGTGACAAAGAAAGCCCCCATC 206

QY 2481 TCTGATTGTTCCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA 2540  
Db 205 TCTGATTGTTCCACTGGAAAGCCAGATGACAAAGAAAGAAATTCACAGGGAAGAAAA 146

QY 2541 AGAGAACAGTAGAATGGAGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAGTGTACT 2600  
Db 145 AGAGAACAGTAGAATGGAGAAAAATGCTGAAAAATCACATAGGCGTTACTGAAGTGTACT 86

QY 2601 TGGAAAGAGCTGCAGCATACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAA 2660

Db 85 TGAAGAAAGCTGCAGCATACAGATAGTTACTTGGGCTTTTCCCATGGGAGAAAA 26  
QY 2661 AAAGTATTTCTAGATCTTCTCGAC 2685  
Db 25 AAAGTATTTCCAAAGATCTTCTCGAC 1

RESULT 8  
AAK15732/c  
ID AAK15732 standard; DNA; 1125 BP.  
AC AAK15732;  
XX 05-NOV-2001 (first entry)  
DT Human brain expressed single exon probe SEQ ID NO: 15723.  
DE Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX Homo sapiens.  
OS WO200157275-A2.  
PN 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US00667.  
XX 04-FEB-2000; 2000US-0180312.  
PF 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human brains -  
PT Example 4; SEQ ID NO: 15723; 650pp + Sequence Listing; English.  
PS The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.  
XX Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 Other;  
SQ

Query Match 30.6%; Score 1101.8; DB 22; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-282;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1581 AGATCAATTTTCATGAATGTTATTAAGTGTATCTCTCCCAACACAGACTCAGTATATAT 1640  
Db 1105 AGATCAATTTTCATGAATGTTATTAAGTGTATCTCTCCCAACACAGACTCAGTATATAT 1046

QY 1641 TCCAAAGGTGATGCTGCTGCTTATTCAGCTTTCGCAAGTAGCCAAAGAGAGGTGA 1700  
Db 1045 TCCAAAGGTGATGCTGCTGCTTATTCAGCTTTCGCAAGTAGCCAAAGAGAGGTGA 986

QY 1701 AGTGCTATAGTACAAATCCAAATATTCGACATGCTTCTATTCGCAACAGAGTGA 1760

Db	985	AGGTGCCTATAGTGACAAATCCAAATAATTCGACATGCTCTATTGCGCAACAGCTGGAAAC	926
QY	1761	CATCCACCTCATATGTCACAGTGGAAATGCAATGCCACCACAATACATTTTAAATCTCAGGTT	1820
Db	925	CATCCACCTCATATGTCACAGTGGAAATGCAATGCCACCACAATACATTTTAAATCTCAGGTT	866
QY	1821	TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGGTGGACACAAGGGA	1880
Db	865	TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGGTGGACACAAGGGA	806
QY	1881	GGGACCAAACTGAATTCACGCCCCAGAAAGGTTACGAAAATTTAGTTAGTCCCATAAAC	1940
Db	805	GGGACCAAACTGAATTCACGCCCCAGAAAGGTTACGAAAATTTAGTTAGTCCCATAAAC	746
QY	1941	ACTTCTTCCAGAGCGGAAATCCTTTTCAGGATATTCCTCAAGAAACACGTTCCCGAA	2000
Db	745	ACTTCTTCCAGAGCGGAAATCCTTTTCAGGATATTCCTCAAGAAACACGTTCCCGAA	686
QY	2001	GTTTAAAGACATGATGTTTAACTCAACAAGGAGAGCCAGGAAGAGGTGAAAATTTCCCTC	2060
Db	685	GTTTAAAGACATGATGTTTAACTCAACAAGGAGAGCCAGGAAGAGGTGAAAATTTCCCTC	626
QY	2061	GGTAAATATTTCACCTCCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCA	2120
Db	625	GGTAAATATTTCACCTCCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCA	566
QY	2121	ACTGGAACATGGAGACATCACTTTGAAGAGATACAAATTTGCGAAGTCAGCCTTGCTGAG	2180
Db	565	ACTGGAACATGGAGACATCACTTTGAAGAGATACAAATTTGCGAAGTCAGCCTTGCTGAG	506
QY	2181	ATCATTTCTGATGAATCAGCATGCTAAAATAAAAAATCAAGCTATAATAACAGATGA	2240
Db	505	ATCATTTCTGATGAATCAGCATGCTAAAATAAAAAATCAAGCTATAATAACAGATGA	446
QY	2241	AACAAATGACAGTTTGGTGGCTCCACAGAAACACAGTTCTATAAAGCATCTTGCCAAA	2300
Db	445	AACAAATGACAGTTTGGTGGCTCCACAGAAACACAGTTCTATAAAGCATCTTGCCAAA	386
QY	2301	CAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACATTTTCCTGCAAGTGAAGT	2360
Db	385	CAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACATTTTCCTGCAAGTGAAGT	326
QY	2361	GAATGTCATGACCCAGGTCAGATCCACCCCTGGATTTGGAGACCACAGAAATTTAG	2420
Db	325	GAATGTCATGACCCAGGTCAGATCCACCCCTGGATTTGGAGACCACAGAAATTTAG	266
QY	2421	AGTGGAACTCACACCCAAAACCATAGCGGAAATGTCACAAAAGAAAGCCCCCATC	2480
Db	265	AGTGGAACTCACACCCAAAACCATAGCGGAAATGTCACAAAAGAAAGCCCCCATC	206
QY	2481	TCGTATTTGTCACCTGGAAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAA	2540
Db	205	TCGTATTTGTCACCTGGAAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAA	146
QY	2541	AGAGAACAGTGAATGGAGGAAATGCTGAAAATCACATAGGCGTTTACTGAAAGTGTACT	2600
Db	145	AGAGAACAGTGAATGGAGGAAATGCTGAAAATCACATAGGCGTTTACTGAAAGTGTACT	86
QY	2601	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTTCGCAATGGAGAAA	2660
Db	85	TGGAAGAAAGCTGCAGCATTACACAGATAGTTACTTGGGCTTTTTCGCAATGGAGAAA	26
QY	2661	AAAGTATTTCTAGATCTTCGAC	2685
Db	25	AAAGTATTTCTAGATCTTCGAC	1

## RESULT 9

RESULT 9  
AAI22215/C

ID AAI22215 standard; DNA; 1125 BP.

XX

AC

XX

12-OCT-2001 (first entry)

Probe #12148 for gene expression analysis in human cervical cell sample

Probe; human; microarray; gene expression; cervical epithelial cell;

cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

.09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.  
30-JUN-2000; 2000US-0608408

03-AUG-2000; 2000US-0632366.  
21-SEP-2000; 2000US-0234687

27-SEP-2000; 2000US-0236359.  
04-OCT-2000; 2000GB-0024263

(MOLE-%) MOLECULAR DYNAMICS

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

analyzing gene expression in human cervix

Claim 25; SEQ ID No

The present invention relates to human single exon nucleic acids

(SENPs). The present sequence is one such probe. The SENPs are d

microarray, which can be used for measuring

samples derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

### Query Match

Best Local Similarity 99.8%; Pred. No. 1e-282;  
Best Local Similarity 99.8%; Pred. No. 1e-282;  
Best Local Similarity 99.8%; Pred. No. 1e-282;

Matches 1103; Conservative 0; Mismatches

QY 1581 AGATCATTTTCATGAATTGTATAAAGTGATCCTTCTCCCAAACCAGACTACTATATAT 1640

Db 1105 AGATCATTTTCATGAATTGTATAAAGTGATCCTTCTCCCAACCAGACTCATATATAT 1046

QY 1641 TCCAAAAGGTGAATGCCTGCCCTTATTTTCAGCTTTTGCAGAACTAGCCAAABACACCCACTTC 1700

Db  
1045 TCCAAAAGGTGAATGCCTGCCCTTATTTCAGGCTTTGGCAGACGTACCCAAACCAACCACTTCC

QY 1701 AGGTGCCTATAGTGACAATCCAATAATTCCACATCCTCTCTTCCCCCACTTCCTGCGGGGTTCC

Db  
985 AGGTGGCTATAGTGCACATCCATTATTCACATCCTGTTGCGTCTG

OV 1761 CATCCACCTCATATTCACACCTCCATACTATCCTGCGTGGTGTTT

[illegible][illegible]

1000  
950  
900  
850  
800  
750  
700  
650  
600  
550  
500  
450  
400  
350  
300  
250  
200  
150  
100  
50  
0

[illegible]







QY 2181 ATCAATTTCTGATGAACATCAGCATGCTAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 2240  
DB 505 ATCAATTTCTGATGAACATCAGCATGCTAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 446  
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCAATAAGCATCTTCCCAA 2300  
DB 445 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCAATAAGCATCTTCCCAA 386  
QY 2301 CAGCTTAGGAGTCTGAAAGATTGAGAGTTGAGTTTCTCTGAGTGAAGTTAAAGT 2360  
DB 385 CAGCTTAGGAGTCTGAAAGATTGAGAGTTGAGTTTCTCTGAGTGAAGTTAAAGT 326  
QY 2361 GAATGGTCAATGACAGGTTGAGATCCACCTGGACTTGGAGACACAGCAAGATTAG 2420  
DB 325 GAATGGTCAATGACAGGTTGAGATCCACCTGGACTTGGAGACACAGCAAGATTAG 266  
QY 2421 AGTGGAACTCACACCAACCAATAGGCGGAAATGTGACAAAAGAAAGCCGCCATC 2480  
DB 265 AGTGGAACTCACACCAACCAATAGGCGGAAATGTGACAAAAGAAAGCCGCCATC 206  
QY 2481 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAA 2540  
DB 205 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATCACAGGGAAGAAA 146  
QY 2541 AGAAGACAGTAGAATGGAGGAAATGCTGAAATCACATAGCGGTTACTGAAGTTTACT 2600  
DB 145 AGAAGACAGTAGAATGGAGGAAATGCTGAAATCACATAGCGGTTACTGAAGTTTACT 86  
QY 2601 TGAAGAAAGCTGCAGCATTACACAGATAGTTTACTTGGGCTTTTGGCATGGGAGAAA 2660  
DB 85 TGAAGAAAGCTGCAGCATTACACAGATAGTTTACTTGGGCTTTTGGCATGGGAGAAA 26  
QY 2661 AAAGTATTCCTAGATCTCTCGAC 2685  
DB 25 AAAGTATTCCTAGATCTCTCGAC 1

RESULT 11  
AA107914/c  
ID AA107914 standard; DNA; 1125 BP.  
XX  
AC AA107914;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Probe #7905 used to measure gene expression in human breast sample.  
XX  
KW Probe; human; breast disease; breast cancer; development disorder; ss;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
XX 09-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-476286/51.  
XX

PT Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast  
XX  
PS Claim 25; SEQ ID No 7905; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes.  
CC The present sequence is one such probe. The probes are useful for  
CC measuring human gene expression in a human breast sample, where the probe  
CC hybridises at high stringency to a nucleic acid expressed in the human  
CC breast. The probes are useful for predicting, diagnosing, grading,  
CC staging, monitoring and prognosing diseases of the human breast,  
CC particularly those diseases with polygenic aetiology. The diseases  
CC include: breast cancer, disorders of development, inflammatory diseases  
CC of the breast, fibrocystic changes, proliferative breast disease and  
CC non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;

Query Match 30.6%; Score 1101.8; DB 22; Length 1125;  
Best Local Similarity 99.8%; Pred. No. 1e-282;  
Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1581 AGATCATTTTCATGAATTTGATATAAGTATCTTCTCCAAACAGACTCACTATATAT 1640  
DB 1105 AGATCATTTTCATGAATTTGATATAAGTATCTTCTCCAAACAGACTCACTATATAT 1046  
QY 1641 TCCAAAGGTGAATGCCCTTATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGA 1700  
DB 1045 TCCAAAGGTGAATGCCCTTATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGA 986  
QY 1701 AGGTGCCTATAGTGACAAATCCAATAATTCGACATGCTTCTATGCCAACAGTGGAAAAC 1760  
DB 985 AGGTGCCTATAGTGACAAATCCAATAATTCGACATGCTTCTATGCCAACAGTGGAAAAC 926  
QY 1761 CATCCACCTCATATGACAGTGAATGAATGCCACCAATACATTTTAATCTCAGTT 1820  
DB 925 CATCCACCTCATATGACAGTGAATGAATGCCACCAATACATTTTAATCTCAGTT 866  
QY 1821 TCAAAATACAAACGATGAAGATTCAAANTGCAGATACAGTGGAGTGACACAGGGA 1880  
DB 865 TCAAAATACAAACGATGAAGATTCAAANTGCAGATACAGTGGAGTGACACAGGGA 806  
QY 1881 GGGACCAAACTGAATTTCTAGGCCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 1940  
DB 805 GGGACCAAACTGAATTTCTAGGCCCCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 746  
QY 1941 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGTATTTCCCAAGAAAACGCTTCCCGAA 2000  
DB 745 ACTTCTTCCAGAGCGGAAATCCTTTTGGAGTATTTCCCAAGAAAACGCTTCCCGAA 686  
QY 2001 GTTTAAGAGCATGATGTTAACTCAACAGAGAGCCAGAGAGGTGAAAATTTCCCT 2060  
DB 685 GTTTAAGAGCATGATGTTAACTCAACAGAGAGCCAGAGAGGTGAAAATTTCCCT 626  
QY 2061 GGTAAATATTTCACTCCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCA 2120  
DB 625 GGTAAATATTTCACTCCTTCCAAAAGACGCCAGTTGAGTCTCAATACCTTGGATTGCA 566  
QY 2121 ACTGGAACATGGAGACATCACTTTTGAAGGATACAAATTTGTCCCAAGTACGCTTGTCTAG 2180  
DB 565 ACTGGAACATGGAGACATCACTTTTGAAGGATACAAATTTGTCCCAAGTACGCTTGTCTAG 506  
QY 2181 ATCAATTTCTGATGAACATCAGCATGCTAAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 2240  
DB 505 ATCAATTTCTGATGAACATCAGCATGCTAAAATAAATAAATAAATCAAGCTATTAATAACAGATGA 446  
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCAATAAGCATCTTCCCAA 2300  
DB 445 AACAAATGACAGTTTGGTGGCTCCACAGGAAACAGGTTTCAATAAGCATCTTCCCAA 386

QY 2301 CAGCTTAGAGTGTCTGAAAGATTGACAGAGTTGACTTTCTCTGACAGTCACTGTAAGT 2360  
 Db |||||||  
 QY 385 CAGCTTAGAGTGTCTGAAAGATTGACAGAGTTGACTTTCTCTGACAGTCACTGTAAGT 326  
 Db |||||||  
 QY 2361 GAATGTCATGACAGGTTGCAATCCACCTGGACTTGGAGACCCACAGCAAGATTAG 2420  
 Db |||||||  
 QY 325 GAATGTCATGACAGGTTGCAATCCACCTGGACTTGGAGACCCACAGCAAGATTAG 266  
 Db |||||||  
 QY 2421 AGTGGAACTCACACCCCAAAACCATAGCGGGAATGTGACAAAAGAAAGCCCCCATC 2480  
 Db |||||||  
 QY 265 AGTGGAACTCACACCCCAAAACCATAGCGGGAATGTGACAAAAGAAAGCCCCCATC 206  
 Db |||||||  
 QY 2481 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATACACAGGCAAGAAA 2540  
 Db |||||||  
 QY 205 TCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAGAAATACACAGGCAAGAAA 146  
 Db |||||||  
 QY 2541 AGAGACAGTAGAATGAGGAAATGCTGAAATACATAGGCGTTTACTGAAAGTTACT 2600  
 Db |||||||  
 QY 145 AGAGACAGTAGAATGAGGAAATGCTGAAATACATAGGCGTTTACTGAAAGTTACT 86  
 Db |||||||  
 QY 2601 TGAAGAAAGCTGACAGATTACACAGATAGTTACTTGGGCTTTTCCCATGGAGAAA 2660  
 Db |||||||  
 QY 85 TGAAGAAAGCTGACAGATTACACAGATAGTTACTTGGGCTTTTCCCATGGAGAAA 26  
 Db |||||||  
 QY 2661 AAGTATTCTCTAGATCTCTCGAC 2685  
 Db |||||||  
 QY 25 AAGTATTCTCTAGATCTCTCGAC 1  
 Db |||||||

RESULT 12

ABSI5475/C  
 ID ABSI5475 standard; DNA; 1125 BP.  
 XX  
 AC ABSI5475;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 15466.  
 XX  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; Niemann-Pick disease;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagen syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease; open reading frame; ORF.  
 XX

OS Homo sapiens.  
 XX  
 XX WO200186003-A2.  
 PN  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 measure gene expression in human lung samples -

XX  
 PS  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridize at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagen syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe open reading frame of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1125 BP; 269 A; 232 C; 223 G; 401 T; 0 other;  
 Query Match 30.6%; Score 1101.8; DB 24; Length 1125;  
 Best Local Similarity 99.8%; Pred. No. 1e-282;  
 Matches 1103; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1581 AGATCATTTTCATGAATGTGATATAAGTATCTCTCCCAACACAGACTCACTATTAT 1640  
 Db |||||||  
 QY 1105 AGATCATTTTCATGAATGTGATATAAGTATCTCTCCCAACACAGACTCACTATTAT 1046  
 Db |||||||  
 QY 1641 TCCAAAAGGTGAATGCCTGCCTTTATTTTCAGCTTTTCAGAAAGTAGCCAAAAGAGGAGTTGA 1700  
 Db |||||||  
 QY 1045 TCCAAAAGGTGAATGCCTGCCTTTATTTTCAGCTTTTCAGAAAGTAGCCAAAAGAGGAGTTGA 986  
 Db |||||||  
 QY 1701 AGGTGCCTATAGTGACAATCCAAATTAATTCGACATGCTTCTATTCGCAACAGTGGAAAC 1760  
 Db |||||||  
 QY 985 AGGTGCCTATAGTGACAATCCAAATTAATTCGACATGCTTCTATTCGCAACAGTGGAAAC 926  
 Db |||||||  
 QY 1761 CATCCACCTCAATAATGCAGTGGAAATGAATGCCACCAATACATTTTAAATCTCACGTT 1820  
 Db |||||||  
 QY 925 CATCCACCTCAATAATGCAGTGGAAATGAATGCCACCAATACATTTTAAATCTCACGTT 866  
 Db |||||||  
 QY 1821 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAGGGA 1880  
 Db |||||||  
 QY 865 TCAAAATACAAACGATGAAGAGTTCAAAATGCAGATACAGTGGAGTGGACACAGGGA 806  
 Db |||||||  
 QY 1881 GGGACCAAAACTGAATCTACAGCCCGCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 1940  
 Db |||||||  
 QY 805 GGGACCAAAACTGAATCTACAGCCCGCAGAGGGTTACGAAAATTTAGTTAGTCCCATAC 746  
 Db |||||||

Claim 4; SEQ ID No 15466; 634pp; English.

QY 1941 ACTTCTTCCAGAGCGGAATCCTTTTGGAGATATTCCCAAGAAAAAGCGTTCGCCGA 2000  
Db 745 ACTTCTTCCAGAGCGGAATCCTTTTGGAGATATTCCCAAGAAAAAGCGTTCGCCGA 686  
QY 2001 GTTTAAGACATGATGTTAACTCAACAGAGAGAGCCAGGAAGAGGTGAATAATCCCTT 2060  
Db 685 GTTTAAGACATGATGTTAACTCAACAGAGAGAGCCAGGAAGAGGTGAATAATCCCTT 626  
QY 2061 GGTAAATATTCTACTCTTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA 2120  
Db 625 GGTAAATATTCTACTCTTCCAAAGAGCGCCAGTGTGAGTCTCAATACCTTGGATTGCA 566  
QY 2121 ACTGGAACATGGAGACATCACTTTGAAAGATACAAATTTGTCAAGTCAAGCTTGGTAG 2180  
Db 565 ACTGGAACATGGAGACATCACTTTGAAAGATACAAATTTGTCAAGTCAAGCTTGGTAG 506  
QY 2181 ATCATTTCTGATGAACCTACAGCATGCTAAATATAAATAATCAAGCTATATAACAGATCA 2240  
Db 505 ATCATTTCTGATGAACCTACAGCATGCTAAATATAAATAATCAAGCTATATAACAGATCA 446  
QY 2241 AACAAATGACAGTTTGGTGGCTCCACAGAAAAAGAGTTTCATAAAGCATCTTGCCTAAA 2300  
Db 445 AACAAATGACAGTTTGGTGGCTCCACAGAAAAAGAGTTTCATAAAGCATCTTGCCTAAA 386  
QY 2301 CAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACTTTTCTGCAAGTGAAGT 2360  
Db 385 CAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACTTTTCTGCAAGTGAAGT 326  
QY 2361 GAATGTCATGACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 2420  
Db 325 GAATGTCATGACAGGTCAGAAATCCACCCCTGGACTTGGAGACACAGCAAGATTAG 266  
QY 2421 AGTGGAACTTCACACCCCAAAACCATTAGCGGAATGTGACAAAGAAAAAGCCCATC 2480  
Db 265 AGTGGAACTTCACACCCCAAAACCATTAGCGGAATGTGACAAAGAAAAAGCCCATC 206  
QY 2481 TCTGATTGTTCCTACTGGAAGCCAGATGACAAAGAAAAAGAAATCACAGGGAAGAAAA 2540  
Db 205 TCTGATTGTTCCTACTGGAAGCCAGATGACAAAGAAAAAGAAATCACAGGGAAGAAAA 146  
QY 2541 AGAAGACAGTAGAATGGAGAAAAATGCTGAAATACATAGCGGTTACTGAAGTGTACT 2600  
Db 145 AGAAGACAGTAGAATGGAGAAAAATGCTGAAATACATAGCGGTTACTGAAGTGTACT 86  
QY 2601 TGAAGAAAGCTGCACGATTACACAGATAGTTACTTTGGCTTTTGGCCATGGAGAAAAA 2660  
Db 85 TGAAGAAAGCTGCACGATTACACAGATAGTTACTTTGGCTTTTGGCCATGGAGAAAAA 26  
QY 2661 AAGATATTCTAGATCTTCTCGAC 2685  
Db 25 AAGATATTCTAGATCTTCTCGAC 1

RESULT 13  
AAV86143  
ID AAV86143 standard; cDNA; 526 BP.  
AC AAV86143;  
XX AAV86143;  
XX AAV86143;  
DT 27-APR-1999 (first entry)  
DE EST clone H545.  
XX  
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
OS Homo sapiens.  
XX  
PN WO9845435-A2.  
XX

PD 15-OCT-1998.  
XX  
PF 10-APR-1998; 98WO-US06954.  
XX  
PR 10-APR-1997; 97US-0835913.  
XX  
PA (GEM ) GENETICS INST INC.  
XX  
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
PI Racie LA, Spaulding V, Treacy M;  
XX  
DR WPI; 1999-070076/06.  
XX  
PT New polynucleotides encoding human secreted proteins - derived from  
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
PT ovary, pituitary, retina and colon cDNA libraries  
XX  
PS Claim 1; Page 135; 633pp; English.  
XX  
CC This sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene  
CC therapy.  
XX  
SQ Sequence 526 BP; 176 A; 112 C; 100 G; 138 T; 0 other;

Query Match 13.9%; Score 501; DB 20; Length 526;  
Best Local Similarity 99.08; Pred. No. 5.9e-123;  
Matches 504; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 2894 AAGAAATTTGACAGAGCGTCATTTCACAAAGTGCGCCATCTGAGGATATGAGTTGGCTT 2953  
Db 12 AAGAGCGCTACAAGAGCGTCATTTCACAAAGTGCGCCATCTGAGGATATGAGTTGGCTT 71  
QY 2954 TCTCTTATTTTATTTATCTCATGAGTGCAGTGCAGGCACCTGAATATATCTCAAGTCTTTG 3013  
Db 72 TCTCTTATTTTATTTATCTCATGAGTGCAGTGCAGGCACCTGAATATATCTCAAGTCTTTG 131  
QY 3014 ATGAAGTTGATACAGATCAATCTGGTCTGTCTGACAGAGAAATCCGACACTGGCTA 3073  
Db 132 ATGAAGTTGATACAGATCAATCTGGTCTGTCTGACAGAGAAATCCGACACTGGCTA 191  
QY 3074 CCAGAAATTCAGAACTGCCGTTAAGTTTGCAGGATTTGACAGGTCTGGAACACATGCTAA 3133  
Db 192 CCAGAAATTCAGAACTGCCGTTAAGTTTGCAGGATTTGACAGGTCTGGAACACATGCTAA 251  
QY 3134 TAAATTCGTCACAAATGCTTCTCTGATATACGAGCTAAATAATATTTCCACCAACTC 3193  
Db 252 TAAATTCGTCACAAATGCTTCTCTGATATACGAGCTAAATAATATTTCCACCAACTC 311  
QY 3194 AGGAATCCTACTATGATCCCAACCTGCCAGCGTCACTAAAAGTCTAGTAACAACTGTA 3253  
Db 312 AGGAATCCTACTATGATCCCAACCTGCCAGCGTCACTAAAAGTCTAGTAACAACTGTA 371  
QY 3254 AACCAAGTACAGCAAAATCCACAAAGCATATAGGACAAAAACAAATATAGTTTGA 3313  
Db 372 AACCAAGTACAGCAAAATCCACAAAGCATATAGGACAAAAACAAATATAGTTTGA 431  
QY 3314 TCATGGGAGAGAGAATCGCTTTTAAATGATTCGTACCAAGCTTTCTCATGTGGTTG 3373  
Db 432 TCATGGGAGAGAGAATCGCTTTTAAATGATTCGTACCAAGCTTTCTCATGTGGTTG 491  
QY 3374 GCCAGTTGGATGACATAGAAAAAACCCCT 3402

DB 492 GCCAGTGGATGACATAGAAAAACCT 520  
 |||||  
 RESULT 14  
 AAA43989  
 ID AAA43989 standard; cDNA; 537 BP.  
 XX  
 AC AAA43989;  
 XX  
 DT 21-AUG-2000 (first entry)  
 XX  
 DE Mouse secreted expressed sequence tag SEQ ID NO:564.  
 XX  
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;  
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;  
 KW antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;  
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KW central nervous system disorder; Alzheimer's disease; stroke;  
 KW parkinson's disease; Huntington's disease; coagulation disorder;  
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KW tumour; infection; depression; psoriasis; ss.  
 OS  
 XS Mus musculus.  
 XX  
 PN WO200021991-A1.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 15-OCT-1999; 99WO-US24206.  
 XX  
 PR 15-OCT-1998; 98US-0104436.  
 XX  
 PA (GENY ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Bowman MR;  
 PI  
 DR WPI; 2000-317938/27.  
 XX  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (SESTs), useful for treating various disorders -  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 XX  
 PS Claim 1; Page 352; 803pp; English.  
 XX  
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed  
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat  
 CC tissue sources. The SESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
 CC antiasthmatic; vulnary; antitumor; osteoprotective; neuroprotective;  
 CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
 CC therapy and in vaccines. The SESTs are useful as probes for the  
 CC identification and isolation of full-length cDNAs and genomic DNA  
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's, disease, stroke), coagulation  
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and

CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
 CC in the exemplification of the present invention.  
 XX  
 SQ Sequence 537 BP; 168 A; 133 C; 105 G; 129 T; 2 other;  
 Query Match 11.1%; Score 398.4; DB 21; Length 537;  
 Best Local Similarity 87.5%; Pred. No. 1.2e-95;  
 Matches 446; Conservative 0; Mismatches 63; Indels 1; Gaps 1;  
 QY 2897 AATTGACAGAGCTCATTTCCAAAAGTGGCGCATTTCTCAGGATATGCGCTTCT 2956  
 DB 21 AATTGACAGAGA-TTCATTTCAAGAGTGGCTCCTCCTCAGGACATGCGCTTCT 79  
 QY 2957 CTTATTTTATTATTCATGAGTGCAGTGCAGCCACTGAATATATCATCAAGTCTTTGATG 3016  
 DB 80 CCTACTTTTATTAACCTCATGAGTGCAGTTCAGCCCTCAATATTTCCCAAGTCTTTTCATG 139  
 QY 3017 AAGTTGATACAGATCAATCTGTTGTTCTGTCGACAGAGAAATCCGACACTGGCTACCA 3076  
 DB 140 AAGTAGACACAGACCAATCTGTTGTTCTGTCGATAGGAAATCCGACACTGGCCACGA 199  
 QY 3077 GAATTCCAGCAACTGCCGTTAAGTTTCAGAGTTTGACAGTCTGGAACACACACTGCTAATAA 3136  
 DB 200 GAATTCACGACCTTACCTTTAAGCTTCAGAGTTTGACAGTTCGGAACACACACTGTTAATAA 259  
 QY 3137 ATTGCTCAAAATGCTTCTGCTGATATCAGCAGCTAATAATATTCCACCAACTCAGG 3196  
 DB 260 ATTGCTCAAAATGCTTCTGCTGATATCAGTCACTCACTCAACACATCCCACTGCTCAGG 319  
 QY 3197 AATCTTACTATGATCCGACCTGCCCGTCACTAAAGTCTAGTAACAACTGTAAC 3256  
 DB 320 AAGCATACTACGACCCCACTGCCCTCCGTCCTTCTGACCAACTGTGAAGC 379  
 QY 3257 CAGTAAGTCAAAATCCCAAGACATATAAGGACAAACAAATATAGTGTGGAATCA 3316  
 DB 380 CAGTAAGTCAAGATCCCAAGACCTATAAAGACAAACAAATACAGTGTGNNATCA 439  
 QY 3317 TGGGAGAGAGAAATGCGTTTAAATGATTCGTACCAACGTTTCTCATGTGGTGGCC 3376  
 DB 440 TGGGAGAGAGAAATGCGTTTCAAGATGATAGCAACCAATGTTCTCATGTGGTGGCC 499  
 QY 3377 AGTTGGATGACATAGAAAAACCCCTAGGA 3406  
 DB 500 AGTTGGATGACATAGAAAAACCCCTAGGA 529  
 RESULT 15  
 ABQ56917  
 ID ABQ56917 standard; cDNA; 572 BP.  
 XX  
 AC ABQ56917;  
 XX  
 DT 02-AUG-2002 (first entry)  
 XX  
 DE Human colon cancer related nucleotide sequence SEQ ID NO:612.  
 XX  
 KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200229086-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 02-OCT-2001; 2001WO-US30732.  
 XX  
 PR 02-OCT-2000; 2000US-237271P.  
 XX  
 PA (FARB ) BAYER CORP.  
 XX  
 PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
 PI Thiagalingam A, Lewis ME;

```
XX WPI; 2002-426115/45.
XX
XX PT New isolated nucleic acid that is differentially expressed in cancer
XX PT tissues useful for determining the presence of colon cancer in a cell
XX PT or tissue type, and in antisense therapy
XX
XX Claim 1; Fig 1; 796pp; English.
XX
XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
XX expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
XX encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
XX used in antisense therapy. An antibody immunoreactive with a polypeptide
XX encoded by (I) is useful for detecting cancer in a patient sample, and
XX for detecting the presence or absence of a polynucleotide encoded by a
XX nucleic acid which hybridizes to (I) in a cell. A probe/primer derived
XX from (I) can be used for determining the presence of a nucleic acid which
XX hybridizes to (I), and for determining the phenotype of cells in a sample
XX of cells from a patient. (I) is useful for determining the presence of
XX colon cancer in a cell or tissue type, for determining the presence of
XX state of other type of cancer, in antisense therapy, to generate
XX macroarrays on a solid surface, to identify a chromosome on which the
XX corresponding gene resides, and in tissue profiling, forensics, genetic
XX analysis, mapping and diagnostic applications. (I) can be used to raise
XX antibodies, and to screen for peptide analogues and antagonists.
XX
XX Sequence 572 BP; 208 A; 124 C; 111 G; 118 T; 11 other;
```

Query Match 9.4%; Score 339; DB 24; Length 572;  
Best Local Similarity 88.6%; Pred. No. 8.1e-80;  
Matches 435; Conservative 0; Mismatches 47; Indels 9; Gaps 6;

```
QY 2035 GCCCAGGAGAGGTGAATTCCTCCCTGGTAAATATTTCACCTCTCCAAAAGACGCCCAG 2094
Db 3 GCGGGGGAAGAGGTGAATTCCTCCCTGGTAAATATTTCACCTCTCCAAAAGACGCCCAG 62

QY 2095 TTGAGTCTCAATACCTTGGATTGCAACTGGAACATGGAGACATCACCTTTGAAAGGATAC 2154
Db 63 TTGAGTCTCAATACCTTGGATTGCAACTGGAACATGGAGACATCACCTTTGAAAGGATAC 122

QY 2155 AATTTGTCCAAGTCAGCCTTGCTGAGATCATTTCTGATGAACCTCACAGCATGCTAAATA 2214
Db 123 AATTTGTCCAAGTCAGCCTTGCTGAGATCATTTCTGATGAACCTCACAGCATGCTAAATA 182

QY 2215 AAAAATCAAGCTATATACAGATCAAAACAAATGACAGTCTTGGTGGCTCCACAGAGAAA 2274
Db 183 AAAAATCAAGCTATATACAGATCAAAACAAATGACAGTCTTGGTGGCTCCACAGAGAAA 242

QY 2275 CAGGTTTCATAAAAGCATCTTG-CCAAACAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTT 2333
Db 243 CAGGTTTCATAAAAGCATCTTGCCCAACACAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTT 302

QY 2334 GACTTTTCTCGAGTGAGTGTAAGAGTGAATGGTCATGACCGGTCAGAAATCCACCCCT 2393
Db 303 GACTTTT-CTGCAGTGAGTGTAAG--GTGAATGGCATGACCA-GGTCAGAATCCACCCCT 358

QY 2394 GCACCTGGAGACACAGCAAGATTTAGAGTGAAGTCAACACACCCCAAAAACCATAGGCGG 2453
Db 359 GCCTTGGAGACCCA---CAGATTATAGTGAAGTCAACACACCC-AAAAACCATAGGCGG 414

QY 2454 AAATGTGACAAAAGAAAAGCCCATCTCTGATTGTTCCACTGGAAAGCCAGATGACAAA 2513
Db 415 AAATGTGACAAAAGAAAAGCCCATCTCTGATTGTTCCACTGGAAAGCCAGATGACAAA 474

QY 2514 AGAAAGAAAA 2524
Db 475 AAAAAAAAAA 485
```

Search completed: April 12, 2003, 22:58:32  
Job time : 539 secs



GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:48:45 ; Search time 85 Seconds  
(without alignments)  
12988.663 Million cell updates/sec

Title: US-10-023-888-1

Perfect score: 3600

Sequence: 1 atggagacagacacactctct.....atcgagacaaattgaagtag 3600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	71.4	2.0	5064	4 US-08-936-107A-8
2	67	1.9	7218	1 US-08-232-463-14
3	63	1.8	417	1 US-08-318-970B-46
4	62.4	1.7	660	2 US-08-302-516-1
5	61.4	1.7	393	1 US-08-137-117D-24
6	61.4	1.7	393	1 US-08-436-717-24
7	61.4	1.7	396	2 US-08-379-057-15
8	61.4	1.7	438	2 US-08-653-402B-11
9	58.2	1.6	393	1 US-08-137-117D-32
10	58.2	1.6	393	1 US-08-436-717-32
11	58.2	1.6	393	2 US-08-621-751A-9
12	58.2	1.6	393	4 US-08-579-378A-13
13	58.2	1.6	396	2 US-08-579-378A-17
14	58.2	1.6	396	2 US-08-483-636-1
15	58.2	1.6	396	2 US-08-483-632-1
16	56.6	1.6	122	1 US-07-634-278-78
17	56.6	1.6	122	1 US-08-477-728-78
18	56.6	1.6	122	1 US-08-474-040-78
19	56.6	1.6	122	1 US-08-487-200-78
20	56.6	1.6	122	4 US-08-484-537-78
21	56.6	1.6	393	1 US-07-634-278-66
22	56.6	1.6	393	1 US-08-477-728-66
23	56.6	1.6	393	1 US-08-474-040-66
24	56.6	1.6	393	1 US-08-487-200-66
25	56.6	1.6	393	4 US-08-484-537-66
26	56.6	1.6	428	3 US-08-589-939-4
27	56.6	1.6	8068	4 US-09-301-593-35

28 55.6 1.5 434 5 PCT-US93-11611-8 Sequence 8, Appli  
29 53.4 1.5 394 3 US-08-836-561-24 Sequence 24, Appli  
30 53 1.5 690 4 US-08-348-548-5 Sequence 5, Appli  
31 53 1.5 690 5 PCT-US95-15716-5 Sequence 5, Appli  
32 46.4 1.3 59 3 US-08-843-409-13 Sequence 13, Appli  
33 45.4 1.3 393 2 US-08-621-751A-13 Sequence 13, Appli  
34 44.6 1.2 4949 3 US-09-138-024-22 Sequence 22, Appli  
35 44.6 1.2 4949 4 US-09-404-066-22 Sequence 22, Appli  
36 44.4 1.2 115 1 US-07-634-278-106 Sequence 106, App  
37 44.4 1.2 115 1 US-08-477-728-106 Sequence 106, App  
38 44.4 1.2 115 1 US-08-474-040-106 Sequence 106, App  
39 44.4 1.2 115 1 US-08-487-200-106 Sequence 106, App  
40 44.4 1.2 115 4 US-08-484-537-106 Sequence 106, App  
41 44.4 1.2 117 1 US-08-458-516-18 Sequence 18, Appli  
42 44.4 1.2 120 1 US-07-634-278-26 Sequence 26, Appli  
43 44.4 1.2 120 1 US-08-477-728-26 Sequence 26, Appli  
44 44.4 1.2 120 1 US-08-474-040-26 Sequence 26, Appli  
45 44.4 1.2 120 1 US-08-487-200-26 Sequence 26, Appli

## ALIGNMENTS

## RESULT 1

US-08-936-107A-8

; Sequence 8, Application US/08936107A

; Patent No. 6403306

; GENERAL INFORMATION:

; APPLICANT: Stephens, David S.

; APPLICANT: Swatley, John S.

; TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences

; TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: US

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/936,107A

; FILING DATE: 23-SEP-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/827,622

; FILING DATE: 09-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Caruthers, Jennie M.

; REGISTRATION NUMBER: 34,464

; REFERENCE/DOCKET NUMBER: 77-97

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080

; TELEFAX: (303) 499-8089

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5064 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: not relevant

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 479..1597

; FEATURE:

; NAME/KEY: CDS

LOCATION: 1598..3236  
FEATURE: CDS  
NAME/KEY: 3309..4052  
LOCATION: 3309..4052  
FEATURE: CDS  
NAME/KEY: 4054..4917  
LOCATION: 4054..4917  
US-08-936-107A-8

Query Match 2.0%; Score 71.4; DB 4; Length 5064;  
Best Local Similarity 55.4%; Pred. No. 6.3e-10;  
Matches 160; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

QY 945 TRACGAAGAACTGAGTACTATCGGATCTATCGAGAGGATGACACCATGGTTCGAA 1004  
DB 2417 TAGAGATGAATAAATTCGCATAGCTCTTGGGAAATGAGTGGATCTTCATTCGAA 2476

QY 1005 TATTTTCATGTACCAACGGGAGATTCCTCGCTGGAACCTTGACATCCCTCGAGT 1064  
DB 2477 AATTTTATTTGCTCTAAATTTGCTCTCCCGAGCATGGCTAGATTTAAATACCTTAAAT 2536

QY 1065 GACAAATAGTAACACACGAGATGTTTTTCGAAATTTGAGCCACTTGCCTTACCTTTAGTTC 1124  
DB 2537 TCAATGGTATATACGAAGAAATATGACAA---AGTGGCTTCTCTACTTTTAGCTC 2593

QY 1125 ACCTGCTATTAAGTACGCTTCATCGCATCGAAGGCTGTCCCGAAGTATTATTACCT 1184  
DB 2594 ACATGCTATTAAGAACGAGCTTGACCATATACCGAAGATAGTAATTTATTATACAG 2653

QY 1185 AAATGATGATGTCTGTTGGGAGGATGCTGGCCAGATGATTTTAC 1233  
DB 2654 CAATGAGACTTCTTAACTAAACCAATGATAAAGACAATTTCTTC 2702

RESULT 2  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: ptz9pt-Fls  
US-08-232-463-14

Query Match 1.9%; Score 67; DB 1; Length 7218;  
Best Local Similarity 4.1%; Pred. No. 1.3e-08;  
Matches 16; Conservative

QY 2381 AGAATCCACCCCTGGACTTGGAGACACACGAAGATTAGAGTGGAACTCACCACCAA 2440  
DB 1446 AGAATTTGGTACRR 1387

QY 2441 AAACCATAGCGGGAATGTGACAAAGAAAGCCCCCATCTCTGATTTGTTCCACTGAAA 2500  
DB 1386 RRR 1327

QY 2501 GCCAGATGACAAAGAAAGAAATACAGGGAAGAAAGAAAGACAGATAGATGAGG 2560  
DB 1326 RRR 1267

QY 2561 AAAATGCTGAAATACATAGGCGTTACTGAAGTGTACTTGGAGAAAGCTGCAGCAT 2620  
DB 1266 RRR 1207

QY 2621 ACACAGATAGTACTTGGGCTTTTGGCATGGGAAAGAAAGATTTCTAGATCTTC 2680  
DB 1206 RRR 1147

QY 2681 TCGACGAAGAGAGTCAATTGAAGACACAATTGGCTTACTTCACTAGCAGAAATAGAG 2740  
DB 1146 RRR 1087

QY 2741 CCAGATACAGAGATACATTTGCGAGATTCCTC 2775  
DB 1086 RRR 1052

RESULT 3  
US-08-318-970B-46  
Sequence 46, Application US/08318970B  
Patent No. 5589573  
GENERAL INFORMATION:  
APPLICANT: Hideaki HAGIWARA, et al.  
TITLE OF INVENTION: AMINO ACID SEQUENCES OF ANTI-IDIOTYPIC  
TITLE OF INVENTION: ANTIBODIES AGAINST ANTI-CANCER HUMAN MONOCLONAL ANTIBODY  
TITLE OF INVENTION: AND DNA BASE SEQUENCES ENCODING THOSE SEQUENCES  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Office of Sherman and Shalloway  
STREET: 413 N. Washington Street  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: Dell System 210; Intel 80 285 Microprocessor  
OPERATING SYSTEM: MS DOS 3.3  
SOFTWARE: Word Perfect, Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/318,970B  
FILING DATE: October 6, 1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Richard A. Steinberg  
REGISTRATION NUMBER: 26,588  
REFERENCE/DOCKET NUMBER: S-2371



RESULT 4  
 US-08-902-516-1  
 ; Sequence 1, Application US/08902516  
 ; Patent No. 5891432  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Soo Hoo, William  
 ; TITLE OF INVENTION: MEMBRANE-BOUND CYTOKINE COMPOSITIONS  
 ; TITLE OF INVENTION: COMBINING GM-CSF AND METHODS OF MODULATING AN IMMUNE  
 ; TITLE OF INVENTION: RESPONSE USING SAME  
 ; NUMBER OF SEQUENCES: 50  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CAMPBELL & FLORES, LLP  
 ; STREET: 4370 La Jolla Village Drive, Suite 700  
 ; CITY: San Diego  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 92121  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/902.516  
 ; FILING DATE: 29-JUL-1997  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Campbell, Cathryn A.  
 ; REGISTRATION NUMBER: 31,815  
 ; REFERENCE/DOCKET NUMBER: P-IM 2442  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619)535-9001  
 ; TELEFAX: (619)535-8949  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 660 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: both  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..660

\_\_\_\_\_

100



01 GRAC 03

RESULT 11  
US-08-621-751A-9  
; Sequence 9, Application US/08621751A  
; Patent No. 5882644  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Chung N.  
; APPLICANT: Landolfi, Nicholas F.  
; APPLICANT: Martin, Ulrich  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES SPECIFIC FOR THE  
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR RECEPTOR  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: MORRISON & FOERSTER LLP  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/621,751A  
; FILING DATE: 22-MAR-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lehnhardt, Susan K.  
; REGISTRATION NUMBER: 33,943  
; REFERENCE/DOCKET NUMBER: 321152000100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 813-5600  
; TELEFAX: (650) 494-0792  
; TELEX: 706141 MRSN FOERS SFO  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 393 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..393  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 61..393  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: group(130..174, 220..240, 337..360)  
; OTHER INFORMATION: /note="Complementarity Determining  
; OTHER INFORMATION: Regions(CDR-1, CDR-2 and CDR-3)"  
US-08-621-751A-9

Query Match 1.6%; Score 58.2; DB 2; Length 393;  
Best Local Similarity 95.2%; Pred. No. 9e-07;  
Matches 60; Conservative 0; Mismatches 3; Indels

1	ATGGAGAGACACACATCCCTGCTATGGGTACGCTGCTCTGGGTTCCAGGTTCCACAGGT	60
QY		
1	ATGGAGAGAGAAACACTCCTGCTATGGGTGCTGCTGCTCTGGGTTCCAGGTTCCACAGGT	60
Db		
QY	61 GAC 63	
Db	61 GAC 63	
RESULT 12		
US-08-579-378A-13		

/ TITLE OF INVENTION: Humanized Antibodies Reactive with  
 / TITLE OF INVENTION: L-Selectin  
 / NUMBER OF SEQUENCES: 20  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Townsend and Townsend and Crew  
 / STREET: One Market Plaza, Steuart Tower, Suite 2000  
 / CITY: San Francisco  
 / STATE: California  
 / COUNTRY: USA  
 / ZIP: 94105  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: PatentIn Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/579,378A  
 / FILING DATE: 27-DEC-1995  
 / CLASSIFICATION: 424  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/160,074  
 / FILING DATE: 30-NOV-1993  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 07/983,946  
 / FILING DATE: 01-DEC-1992  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: EP 95112895.8  
 / FILING DATE: 17-AUG-1995  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: EP 95114696.8  
 / FILING DATE: 19-SEP-1995  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Liebescheutz, Joe O.  
 / REGISTRATION NUMBER: 37,505  
 / REFERENCE/DOCKET NUMBER: 11823-002220  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 415-326-2400  
 / TELEFAX: 415-326-2422  
 / INFORMATION FOR SEQ ID NO: 17:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 393 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: cDNA  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: 1..393  
 / PS-08-579-378A-17

QY	61	GAC	63
Db	61	GAC	63

RESULT 14  
US-08-483-636-1  
; Sequence 1, Application US/08483636  
; Patent No. 5914110  
; GENERAL INFORMATION:  
; APPLICANT: Holmes, Stephen D.  
; APPLICANT: Gross, Mitchell S.  
; APPLICANT: Sylvester, Daniel R.  
; TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
; TITLE OF INVENTION: Treatment of IL4 Mediated Disorders

NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESS: Intellectual Property  
STREET: P.O. Box 1539 / UW2220  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,636  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117366  
FILING DATE: 07-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/136783  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US/94/10308  
FILING DATE: 07-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P50186-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5024  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 396 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..396  
US-08-483-636-1

Query Match 1.6%; Score 58.2; DB 2; Length 396;  
Best Local Similarity 95.2%; Pred. No. 9e-07;  
Matches 60; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGAGACAGACACTCTGCTATGGTACTGCTCTGGTTCCAGGTTCCACTGGT 60  
|||||  
Db 1 ATGGAGACAGACAAATCTGCTATGGTACTGCTCTGGTTCCAGGTTCCACTGGT 60  
QY 61 GAC 63  
|||  
Db 61 GAC 63

RESULT 15  
US-08-483-632-1  
Sequence 1, Application US/08483632  
Patent No. 5928904  
GENERAL INFORMATION:  
APPLICANT: Holmes, Stephen D.  
APPLICANT: Gross, Mitchell S.  
APPLICANT: Sylvester, Daniel R.  
TITLE OF INVENTION: Recombinant IL4 Antibodies Useful in  
TITLE OF INVENTION: Treatment of IL4 Mediated Disorders  
NUMBER OF SEQUENCES: 75  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESS: Intellectual Property

STREET: P.O. Box 1539 / UW2220  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,632  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117366  
FILING DATE: 07-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/136783  
FILING DATE: 14-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US/94/10308  
FILING DATE: 07-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Sutton, Jeffrey A.  
REGISTRATION NUMBER: 34,028  
REFERENCE/DOCKET NUMBER: P50186-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5024  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 396 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..396  
US-08-483-632-1

Query Match 1.6%; Score 58.2; DB 2; Length 396;  
Best Local Similarity 95.2%; Pred. No. 9e-07;  
Matches 60; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGGAGACAGACACTCTGCTATGGTACTGCTCTGGTTCCAGGTTCCACTGGT 60  
|||||  
Db 1 ATGGAGACAGACAAATCTGCTATGGTACTGCTCTGGTTCCAGGTTCCACTGGT 60  
QY 61 GAC 63  
|||  
Db 61 GAC 63

Search completed: April 13, 2003, 01:36:21  
Job time : 118 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:48:10 ; Search time 3185 Seconds  
(without alignments)  
18305.731 Million cell updates/sec

Title: US-10-023-888-1

Perfect score: 3600

Sequence: 1 atgagagacagacacactct.....atgagacaaattgaagtag 3600

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*\*  
2: em\_esthum:\*\*  
3: em\_estim:\*\*  
4: em\_estmu:\*\*  
5: em\_estov:\*\*  
6: em\_estpl:\*\*  
7: em\_estro:\*\*  
8: em\_htc:\*\*  
9: gb\_est1:\*\*  
10: gb\_est2:\*\*  
11: gb\_htc:\*\*  
12: gb\_est3:\*\*  
13: gb\_est4:\*\*  
14: gb\_est5:\*\*  
15: em\_estfun:\*\*  
16: em\_estom:\*\*  
17: gb\_gss:\*\*  
18: em\_gss\_hum:\*\*  
19: em\_gss\_inv:\*\*  
20: em\_gss\_pln:\*\*  
21: em\_gss\_vrt:\*\*  
22: em\_gss\_fun:\*\*  
23: em\_gss\_mam:\*\*  
24: em\_gss\_mus:\*\*  
25: em\_gss\_other:\*\*  
26: em\_gss\_pro:\*\*  
27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	771.8	21.4	1061	13	BM450679
2	730.4	20.3	1016	13	BM544697
3	648.6	18.0	984	12	BM250072
4	599.8	16.7	613	12	BG163659
5	589.2	16.4	616	14	BM790504
6	559.6	15.5	623	14	BQ305720

7	559	15.5	567	12	BG291336
8	537.2	15.5	610	14	BQ307539
9	546.2	15.2	654	10	AW578427
10	546.2	15.2	654	10	AW604365
11	544.2	15.1	580	14	BQ306940
12	524.6	14.6	533	9	AI499228
13	514.4	14.3	572	12	BE736417
14	509	14.1	709	12	BF739893
15	507.2	14.1	673	12	BG085937
16	504.8	14.0	672	13	BI103956
17	480	13.3	929	14	BQ720582
18	480	13.3	934	14	BQ719691
19	478	13.3	981	14	BQ553143
20	468.8	13.0	600	9	AL135521
21	464.2	12.9	476	12	BF922316
22	455.4	12.7	693	9	AI664612
23	449.4	12.5	644	10	BE370268
24	440.6	12.2	484	14	BQ344941
25	434.8	12.1	1270	12	BG25753
26	431.2	12.0	596	13	BM364071
27	418.6	11.6	658	14	BQ305706
28	417.4	11.5	528	9	AA204698
29	414.4	11.5	482	12	BF931380
30	411.8	11.4	1037	14	BQ226717
31	397.4	11.0	637	13	BI660298
32	385.6	10.7	470	12	BF931580
33	378.8	10.5	459	12	BF931578
34	370.8	10.3	487	10	BB706869
35	369.6	10.3	1015	12	BF784322
36	350.8	9.7	356	14	N44886
37	343.8	9.6	815	13	BI757655
38	340.8	9.5	506	9	AI882177
39	335.8	9.3	442	12	BF566740
40	325.8	9.0	572	13	BM241944
41	320.2	8.9	550	17	AO549054
42	309.2	8.6	322	12	BF934976
43	303	8.4	570	14	BM798105
44	280.8	7.8	579	12	BE695877
45	280.6	7.8	335	9	AA646238

#### ALIGNMENTS

RESULT 1  
BM450679  
LOCUS  
DEFINITION  
AGENCOURT\_6394543 NIH\_MGC\_67 Homo sapiens cDNA Clone IMAGE:5494379  
5', mRNA sequence.  
ACCESSION  
BM450679  
VERSION  
BM450679.1  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1061)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
AUTHORS  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE  
Unpublished (1999)  
JOURNAL  
Contact: Robert Strausberg, Ph.D.  
COMMENT  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: L14W1219 row: j column: 12  
High quality sequence stop: 637.  
Location/Qualifiers  
1. .1061

BM450679  
AGENCOURT\_6394543 NIH\_MGC\_67 Homo sapiens cDNA Clone IMAGE:5494379  
5', mRNA sequence.

BM450679.1 GI:18499719

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L14W1219 row: j column: 12

High quality sequence stop: 637.

Location/Qualifiers

source

1. .1061

Mon Apr 14 09:20:36 2003

us-10-023-888-1.rst

BASE COUNT	308 a	240 c	234 g	278 t	1 others
ORIGIN					
	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5494379" /clone_lib="NIH_MGC_67" /tissue_type="retinoblastoma" /lab_host="DH10B (phage-resistant)" /notes="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI. Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."				

Query Match	21.4%	Score 771.8	DB 13	Length 1061	
Best Local Similarity	95.9%	Pred. No. 2.6e-170			
Matches 826	Conservative	0	Mismatches 17	Indels 8	Gaps 3
QY	446	CTTCTTTTCATTCTGCAGTGCACATTTTCAATGTTGCAAAACCAAAACCCCTCTACCA	505		
DB	25	CTTCTTTTCATTCTGCAGTGCACATTTTCAATGTTGCAAAACCAAAACCCCTCTACCA	84		
QY	506	ATGCTCTAGTGTGTTTTGACAGTACTAAGAGTGTGAAGATGCCACTCTGCAGTGC	565		
DB	85	ATGCTCTAGTGTGTTTTGACAGTACTAAGAGTGTGAAGATGCCACTCTGCAGTGC	144		
QY	566	TTAAAGGAATAAGCAGACAGACAGTATGGAGGGGCTACTTGACAACAGATAAAGAAGTCC	625		
DB	145	TTAAAGGAATAAGCAGACAGACAGTATGGAGGGGCTACTTGACAACAGATAAAGAAGTCC	204		
QY	626	CTGGATFAGTCTAATGCAAGATTGGCTTTCTGAGTGGATTTCACCAACATTCAGG	685		
DB	205	CTGGATFAGTCTAATGCAAGATTGGCTTTCTGAGTGGATTTCACCAACATTCAGG	264		
QY	686	AACAACTCAACTAAAAACAATAATGCCAGAAAATCTTCTCTAAAGTCAAACTGTTCG	745		
DB	265	AACAACTCAACTAAAAACAATAATGCCAGAAAATCTTCTCTAAAGTCAAACTGTTCG	324		
QY	746	AGTTGTATTTCAGAGGCCAGTGTAGCGCTTCTPAAACTGAATAACCCCAAGGATTTTCAAG	805		
DB	325	AGTTGTATTTCAGAGGCCAGTGTAGCGCTTCTPAAACTGAATAACCCCAAGGATTTTCAAG	384		
QY	806	AATTGAATTAAGCAAACTAAGAGAACATGACATTGATGCAAAAGAACTGACCATTAAGTC	865		
DB	385	AATTGAATTAAGCAAACTAAGAGAACATGACATTGATGCAAAAGAACTGACCATTAAGTC	444		
QY	866	CTGCATATTTATTTATGGGATCTGAGCGGCATCAGCCAGTCTAAAGCAGGATGAAGACATCT	925		
DB	445	CTGCATATTTATTTATGGGATCTGAGCGGCATCAGCCAGTCTAAAGCAGGATGAAGACATCT	504		
QY	926	CTGCCAGTCGTTTTGAAGATAACGAAGAACCTGAGGTACTCATTTGCGATCTATCGAGAGC	985		
DB	505	CTGCCAGTCGTTTTGAAGATAACGAAGAACCTGAGGTACTCATTTGCGATCTATCGAGAGC	564		
QY	986	ATGCACCATTGGTTCGGAAATATTTTCATTGTCCACCAAGGCGAGATCCATCTCGTCTGA	1045		
DB	565	ATGCACCATTGGTTCGGAAATATTTTCATTGTCCACCAAGGCGAGATCCATCTCGTCTGA	624		
QY	1046	ACCTTGACAACTCTCGAGTGACAATAGTAACACACACAGGATGTTTTTCGAAAATTCGAGCC	1105		
DB	625	ACCTTGACAACTCTCGAGTGACAATAGTAACACACACAGGATGTTTTTCGAAAATTCGAGCC	684		
QY	1106	ACTTGCCCTACCTTTAGTTCACCTGCTATTGAAAGTTCAGTTCATCGCATCGAAGGCTGT	1165		
DB	685	ACTTGCCCTACCTTTAGTTCACCTGCTATTGAAAGTTCAGTTCATCGCATCGAAGGCTGT	744		
QY	1166	CCAGAAGTTTATTTACCTAAATGATGTGATGTGCTTTGGG---AAGGATGTCTGGCCAG	1222		
DB	745	CCAGAAGTTTATTTACCTAAATGATGTGATGTGCTTTGGG---AAGGATGTCTGGCCAG	804		
QY	1223	ATGATTTTTACAGTCACTAC---TCCAAAGGCCAGAGGTTTTATTTGACATGGCC---TGTCGC	1277		
DB	805	ATGATTTTTACAGTCACTAC---TCCAAAGGCCAGAGGTTTTATTTGACATGGCC---TGTCGC	864		

QY	1278	AAACTGTGCCGAGGGCTGCC	1298
Dd	865	AAACTGTGCCGAGGGCTGCC	885
RESULT 2			
	BM544697		
	LOCUS	1016 bp	mRNA
	DEFINITION	AGENCOURT_6494522 NIH_MGC_124 Homo sapiens cDNA clone IMAGE:5727417	linear EST 20-FEB-2002
	ACCESSION	BM544697	5', mRNA sequence.
	VERSION	BM544697.1	GI:18776157
	KEYWORDS	EST.	
	SOURCE	human.	
	ORGANISM	Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
	REFERENCE	1 (bases 1 to 1016)	
	AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .	
	TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
	JOURNAL	Unpublished (1999)	
	COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Invitrogen cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Clone distribution: Agencourt Bioscience Corporation found through the I.M.A.G.E. Consortium/LLNL at:	

FEATURES	source	Location/Qualifiers
1. .1016		
/organism="Homo sapiens"		
/db_xref="taxon:9606"		
/clone="IMAGE:5727417"		
/clone_lib="NIH_MGC_124"		
/tissue_type="hippocampus"		
/lab_host="DH10B"		
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: EcorV (destroyed); Site_2: NotI; RNA source male hippocampus, age 27. Library is oligo-dT primed and directionally cloned (EcorV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 0.9-4 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 012."		
BASE COUNT	302 a	226 c
	224 g	260 t
		4 others

BASE COUNT	302 a	226 c	224 g	260 t	4 others
ORIGIN					
Query Match	20.3%;	Score 730.4;	DB 13;	Length 1016;	
Best Local Similarity	99.1%;	Pred. No. 1.3e-160;			
Matches 743;	Conservative	0;	Mismatches 6;	Indels 1;	Gaps 1;
Qy	103	AGCCGAGATCAATACCATGTTTGTTCATTCTCTATAGAGACAATATTGCTGGAAAGTCC	162		
Db	105	AGCCGAGATCAATACCATGTTTGTTCATTCTCTATAGAGACAATATTGCTGGAAAGTCC	164		
Qy	163	TTTTCAGAAATCGGCTTTGTCTGCCCAATGCCGATTGACGTTGTTTACACCTGGGTGAATGCG	222		
Db	165	TTTTCAGAAATCGGCTTTGTCTGCCCAATGCCGATTGACGTTGTTTACACCTGGGTGAATGCG	224		
Qy	223	ACAGATCTTGTAACCTACTCAAGGAACCTACAGCAGGTCAGAGAACACAGATGGAGGAGGAGCAG	282		
Db	225	ACAGATCTTGTAACCTACTCAAGGAACCTACAGCAGGTCAGAGAACACAGATGGAGGAGGAGCAG	284		
Qy	283	AAAGCAATCAGAGAAATCCTTTGGGAAAAACACACGCGAACCTTACTTAAGAAGAGTGCAGAAG	342		
Db	285	AAAGCAATCAGAGAAATCCTTTGGGAAAAACACACGCGAACCTTACTTAAGAAGAGTGCAGAAG	344		
Qy	343	CAGTTAGAGTGTTCCTAACACACTGCATTAAGGTGCCAATGCTTGTCTTGACCCAGCC	402		



Db 345 CAGTTAGAGTGTGTTGCTTAACACACTGCATTAAAGTGGCCAAATGCTTCTCTGGACCCAGCC 404  
 QY 403 CTGCCAGCCAAACATCACCCTGAAGGACCTGCCATCTCTTTATACCTTCTTTTCAATCTGCC 462  
 Db 405 CTGCCAGCCAAACATCACCCTGAAGGACCTGCCATCTCTTTATACCTTCTTTTCAATCTGCC 464  
 QY 463 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCACTTGTGTT 522  
 Db 465 AGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCTCACTTGTGTT 524  
 QY 523 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGGAAATAGCAGA 582  
 Db 525 TTTGACAGTACTAAGGATGTTGAAGATGCCCACTCTGGACTGCTTAAAGGAAATAGCAGA 584  
 QY 583 CACACAGTATGGAGGGCTACTTGACACAGATAAAGAGTCCCTGGATTTAGTGTATG 642  
 Db 585 CACACAGTATGGAGGGCTACTTGACACAGATAAAGAGTCCCTGGATTTAGTGTATG 644  
 QY 643 CAAGATTTGGCTTTCCTGAGTGGATTTCACCAACATTCAGGAACAAATCAACTAAAA 702  
 Db 645 CAAGATTTGGCTTTCCTGAGTGGATTTCACCAACATTCAGGAACAAATCAACTAAAA 704  
 QY 703 ACAAATTCGCAGAAATCTTCTCTAAAGTCAAACTGTTGCAGTTGTATTCAGAGGCC 762  
 Db 705 ACAAATTCGCAGAAATCTTCTCTAAAGTCAAACTGTTGCAGTTGTATTCAGAGGCC 764  
 QY 763 AGTGTAGCGCTTCTAAACCTGAATACCCCAAGGATTTCAAGATTTGAATTAAG-CAAAC 821  
 Db 765 AGTGTAGCGCTTCTAAACCTGAATACCCCAAGGATTTCAAGATTTGAATTAAG-CAAAC 824  
 QY 822 TAAGAAACATGACCATTTGATGGAAAGA 851  
 Db 825 TAAGAAACATGACCATTTGATGGAAAGA 854

RESULT 3  
 BG250072 984 bp mRNA linear EST 13-FEB-2001  
 LOCUS 602362161F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4470615 5',  
 mRNA sequence.  
 ACCESSION BG250072  
 VERSION BG250072.1 GI:12759888  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 984)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLNL0288 row: i column: 16  
 High quality sequence stop: 670.  
 Location/Qualifiers  
 1..984  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4470615"  
 /clone\_lib="NIH\_MGC\_90"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dr primed.  
 Average insert size 1.7 kb. Library enriched for

FEATURES  
source

full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 327 a 201 c 228 g 228 t  
 ORIGIN

Query Match 18.08; Score 648.6; DB 12; Length 984;  
 Best Local Similarity 92.8%; Pred. No. 2.1e-141;  
 Matches 747; Conservative 0; Mismatches 49; Indels 9; Gaps 6;

QY 2269 GAAAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGCAG 2328  
 Db 1 GAAAAACAGGTTTCATAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGCAG 60  
 QY 2329 AGGTGACTTTTCTCGCAGTGAGTGTAAAGTGAATGGTACATGACCGGTCAGAAATCCA 2388  
 Db 61 AGGTGACTTTTCTCGCAGTGAGTGTAAAGTGAATGGTACATGACCGGTCAGAAATCCA 120  
 QY 2389 CCCCTGGACTTGGAGACCCACAGCAAGATTTAGAGTGAAGTCAACCCCAAAACCATTA 2448  
 Db 121 CCCCTGGACTTGGAGACCCACAGCAAGATTTAGAGTGAAGTCAACCCCAAAACCATTA 180  
 QY 2449 GCGGAAATGTGACAAAGAAAGCCCACTCTCTGATTTGTTCCACTGGAAAGCCAGATG 2508  
 Db 181 GCGGAAATGTGACAAAGAAAGCCCACTCTCTGATTTGTTCCACTGGAAAGCCAGATG 240  
 QY 2509 ACAAAGAAAGAAATCAACAGGGAAG-AAAAAGAGAACAGTAGAATGGAGGAAATGC 2567  
 Db 241 ACAAAGAAAGAAATCAACAGGGAAGAAAGAAAGAACAGTAGAATGGAGGAAATGC 300  
 QY 2568 TGAATAATCACATAGCGCTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATTTACACAGA 2627  
 Db 301 TGAATAATCACATAGCGCTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATTTACACAGA 360  
 QY 2628 TAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCTAGATCTTCTCGACGA 2687  
 Db 361 TAGTTACTTGGGCTTTTGGCATGGGAGAAAGAAAGTATTTCTAGATCTTCTCGACGA 420  
 QY 2688 AGAAGATCATTTGAAGACAAATTTGGCTTACTTCACTGTAGCAAGAAATAGAGCCAGATA 2747  
 Db 421 AGAAGATCATTTGAAGACAAATTTGGCTTACTTCACTGTAGCAAGAAATAGAGCCAGATA 480  
 QY 2748 CAAGAGAGATACATTTGCAGATTTCCCTCAGATATGTAATAAATAATCTTAATAGCAAGTT 2807  
 Db 481 ACTAAAGATACATTTGCCGATTTCCCTCAGATATGTAATAAATAATCTTAATAGCAAGTT 540  
 QY 2808 TGGATTCACATCGCGGAAAGTCCCTCAGATGCTCCTCAGATGATTTGACCGGATTTGTTAT 2867  
 Db 541 TGGATTCACATCGCGGAAAGTCCCTCAGATGCTCCTCAGATGATTTGACCGGATTTGTTAT 600  
 QY 2868 GCAAGAACTCAAGATATGTTCCCTGAGAAATTTGCAAGACGCTCATTTCAAAAGTG-C 2926  
 Db 601 GCAAGAACTCAAGATATGTTCCCTGAGAAATTTGCAAGACGCTCATTTCAAAAGTGCC 660  
 QY 2927 GCCATCTGAGGATAT-GCAGTTTGGCTTCTCTTATTTT-TTATATCATGAGTGCAG 2983  
 Db 661 GCCATCTGAGGATATGCGAGTTTGGCTTCTCTTATTTTATGATTTCCATGATTTGCAG 720  
 QY 2984 TGCAGCCACTGAATATA---TCTCAAGTCTTTGATGAAGTTGATA-CAGATCAATCTGTT 3039  
 Db 721 TGCAGCCACTGAATATAATTTCTCAAGTCTTTGCTGGAAGTGTATTACCGGATCCACCTGGG 780  
 QY 3040 GTCTTGTCTCAGACAGAAATCCGAA 3064  
 Db 781 GTCTTGTCTGGAAGAAATTCGAA 805

RESULT 4  
 BG163659  
 LOCUS 60238393F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4446921 5',  
 mRNA sequence.  
 DEFINITION BG163659  
 ACCESSION BG163659  
 VERSION BG163659.1 GI:12670362

BC163659 613 bp mRNA linear EST 06-FEB-2001  
 60238393F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4446921 5',  
 mRNA sequence.  
 BG163659  
 BG163659.1 GI:12670362

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10226 row: n column: 10
High quality sequence stop: 611.
Location/Qualifiers
1..613
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4448921"
/clone_lib="NIH_MGC_89"
/tissue_type="hypermephroma, cell line"
/lab_host="DH10B (phage-resistant)"
Note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 151 a 122 c 166 g 174 t
ORIGIN
Query Match 16.7%; Score 599.8; DB 12; Length 613;
Best Local Similarity 99.7%; Pred. No. 5.6e-130;
Matches 601; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1021 AACGGCAGATCCATCCTGGCTGACCTTGACATCTCCGAGTACATAGTACACAC 1080
|||||
Db 1 AACGGCAGATCCATCCTGGCTGACCTTGACATCTCCGAGTACATAGTACACAC 60
|||||
QY 1081 CAGGATGTTTTGCGAATTTGAGCCACTTGCCTACTCTTTAGTTTCACTCTCTATTGAAAGT 1140
|||||
Db 61 CAGGATGTTTTGCGAATTTGAGCCACTTGCCTACTCTTTAGTTTCACTCTCTATTGAAAGT 120
|||||
QY 1141 CAGGTTATCGCATCGAAGGGCTGTCCAGAAAGTTTATTACCTAAATGATGATGTCATG 1200
|||||
Db 121 CACATTCATCGCATCGAAGGGCTGTCCAGAAAGTTTATTACCTAAATGATGATGTCATG 180
|||||
QY 1201 TTTGGGAAGAGTCTGGCCAGATGATTTTACAGTCACTCCAAAGGCCAGAGCTTAT 1260
|||||
Db 181 TTTGGGAAGAGTCTGGCCAGATGATTTTACAGTCACTCCAAAGGCCAGAGCTTAT 240
|||||
QY 1261 TTGACATGGCTGTGCGCAAACTGTGCGGAGGCTGCCAGAGTTCTCTGGATTAAAGATGCG 1320
|||||
Db 241 TTGACATGGCTGTGCGCAAACTGTGCGGAGGCTGCCAGAGTTCTCTGGATTAAAGATGCG 300
|||||
QY 1321 TATTGTGACAGGCTTGTATAATTCAGCTCGGATGTTGGATGGGGATGCTCTGGA 1380
|||||
Db 301 TATTGTGACAGGCTTGTATAATTCAGCTCGGATGTTGGATGGGGATGCTCTGGA 360
|||||
QY 1381 AACATCGGAGGAGTGCCTATATTGACAGGTTGAGGACTGCGGAGTATTCGAGTTGA 1440
|||||
Db 361 AACATCGGAGGAGTGCCTATATTGACAGGTTGAGGACTGCGGAGTATTCGAGTTGA 420
|||||
QY 1441 CAGCCCTGGCAGTTTGGTGAGGAATAAACAGTGTCTTCTTACTGTAAATCAGGAGTGTGGG 1500
|||||
Db 421 CAGCCCTGGCAGTTTGGTGAGGAATAAACAGTGTCTTCTTACTGTAAATCAGGAGTGTGGG 480
|||||
QY 1501 AATTCCTGGCTCGCTGATAGTTCTGTGACCAAGATGCAATGTCTTGCTCTGGGTTT 1560
|||||

Db 481 AATTCCTGGCTCGCTGATAGTTCTGTGACCAAGATGCAATGTCTTGCTCTGGGTTT 540
|||||
QY 1561 GATGCTGGCGACTGTGGCAAGATCATTTTCATGAATTTGTATAAGTATCTCTCTCCCA 1620
|||||
Db 541 GATGCTGGCGACTGTGGCAAGATCATTTTCATGAATTTGTATAAGTATCTCTCTCCCA 600
|||||
QY 1621 AAC 1623
|||||
Db 601 AAC 603
|||||

RESULT 5
BM790504 616 bp mRNA linear EST 05-MAR-2002
LOCUS K-EST0070348 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-7-F08
DEFINITION 5', mRNA sequence.
ACCESSION BM790504
VERSION BM790504.1 GI:19138736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 616)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 7 row: F column: 08
High quality sequence stop: 616.
Location/Qualifiers
1..616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNU16n1-7-F08"
/clone_lib="S22SNU16n1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/lab_host="SNU-16"
Note="Organ: Stomach; Vector: pT73-Pac; Site_1: EcoRI;
Site_2: NotI; The S22SNU16 library was contributed by the
Soares laboratory and it was constructed as described by the
Bonaldi, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."
BASE COUNT 155 a 117 c 168 g 176 t
ORIGIN
Query Match 16.4%; Score 589.2; DB 14; Length 616;
Best Local Similarity 99.2%; Pred. No. 1.7e-127;
Matches 613; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1139 GTCAGTTTCATCGCATGAAGGGCTGCCAGAAAGTTTATTACCTAAATGATGATGTCGA 1198
|||||
Db 1 GTCACATTTCATCGCATGAAGGGCTGCCAGAAAGTTTATTACCTAAATGATGATGTCGA 60
|||||
QY 1199 TGTTTGGGAAGGATGCTGGCCAGATGATTTTTTACAGTCACTCCAAAGGCCAAGAGTTT 1258
|||||
Db 61 TGTTTGGGAAGGATGCTGGCCAGATGATTTTTTACAGTCACTCCAAAGGCCAAGAGTTT 120
|||||

```

QY 1259 ATTGACATGGCTGTGCGAACTGTGCGAGGCTGCCAGGTCTCCTGGATTAGGATG 1318  
 Db 121 ATTTGACATGGCTGTGCGAACTGTGCGAGGCTGCCAGGTCTCCTGGATTAGGATG 180  
 QY 1319 GCTATTGTGACAGGCTGTGTAATAATTCAGCCTGCGATTGGATGGTGGGATTGCTCTG 1378  
 Db 181 GCTATTGTGACAGGCTGTGTAATAATTCAGCCTGCGATTGGATGGTGGGATTGCTCTG 240  
 QY 1379 GAACAGTGGAGGAGTCTCTATTCAGGAGGTGGAGTACTGGGATATGGAGTTG 1438  
 Db 241 GAACAGTGGAGGAGTCTCTATTCAGGAGGTGGAGTACTGGGATATGGAGTTG 300  
 QY 1439 GACAGCCCTGGAGTGTGCGAGGAATAACAGTGTCTCTACTGTAATCAGGATCTG 1498  
 Db 301 GACAGCCCTGGAGTGTGCGAGGAATAACAGTGTCTCTACTGTAATCAGGATCTG 360  
 QY 1499 CGAATTCCTGGCTGCGTATAGTCTCTGTGACCAAGCATGCAATGCTCTGCTGGGT 1558  
 Db 361 CGAATTCCTGGCTGCGTATAGTCTCTGTGACCAAGCATGCAATGCTCTGCTGGGT 420  
 QY 1559 TTGATGCTGGCAGCTGTGGCAAGATCATTTTCATGAATGTAATAAGTGAFCCTTCTCC 1618  
 Db 421 TTGATGCTGGCAGCTGTGGCAAGATCATTTTCATGAATGTAATAAGTGAFCCTTCTCC 480  
 QY 1619 CAACACAGACTCACTATATTTATTCAAAGGTGAATGCTGCTTATTTACGCTTTGCGAG 1678  
 Db 481 CAACACAGACTCACTATATTTATTCAAAGGTGAATGCTGCTTATTTACGCTTTGCGAG 540  
 QY 1679 AAGTAGCCAAAGAGAGTGAAGTGGCTATAGTACAAATCCAAATATTCGACATGCTT 1738  
 Db 541 AAGTAGCCAAAGAGAGTGAAGTGGCTATAGTACAAATCCAAATATTCGACATGCTT 598  
 QY 1739 CTATTGCCAACAGTGGGA 1756  
 Db 599 CTATTGCCAACAGTGGGA 616

RESULT 6  
 BQ305720/c  
 LOCUS  
 DEFINITION MRO-BT2002-290501-102-g09 BT2002 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BQ305720  
 VERSION BQ305720.1 GI:20835547  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 623)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-BT2002-290501-102-g09&t3=2001-05-29&t4=1)  
 Seq primer: puc 18 forward

High quality sequence stop: 422.  
 Location/Qualifiers  
 1..623  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BT2002"  
 /dev\_stage="Adult"  
 /note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 160 a 128 c 137 g 198 t  
 ORIGIN

Query Match 15.5%; Score 559.6; DB 14; Length 623;  
 Best Local Similarity 97.5%; Pred. No. 1.5e-120;  
 Matches 579; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 2962 TTTTATTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTTGATGAAGTT 3021  
 Db 602 TGTATTATCTCATGAGT-CAGTGCAGCCACTGAATATATCTCAAGTCTTTTGATGAAGTT 544  
 QY 3022 GATACAGATCAATCTGCTGCTCTGTGACAGAGAAATCCGACACGTGGCTACAGAAAT 3081  
 Db 543 GATACAGATCAATCTGCTGCTCTGTGACAGAGAAATCCGACACGTGGCTACAGAAATG 484  
 QY 3082 CACGAACCTGCGTTAACTTTTCAGAGGATTTGACAGGCTGCGAACACATGCTAATAAATTCG 3141  
 Db 483 CACGGGCTGCGTTAACTTTTCAGAGGATTTGACAGGCTGCGAACACATGCTAATAAATTCG 424  
 QY 3142 TCRAAAATGCTTCTGCTGATATCAGCAGCTAAATAATATTCACCACTCAGGAATCC 3201  
 Db 423 TCRAAAATGCTTCTGCTGATATCAGCAGCTAAAGAATATTCGCGCACTCAGGAATCC 364  
 QY 3202 TACTATGATCCCACTCCACCGTCTCACTAAAAGTCTAGTAAACATGTAACACAGTA 3261  
 Db 363 TACTATGATCCCACTCCACCGTCTCACTAAAAGTCTAGTAAACATGTAACACAGTA 304  
 QY 3262 ACTGACAAAATCCACAAGCATATAGGACAAAACAAATATAGGTTTGAATCATGGGA 3321  
 Db 303 ACTGACAAAATCCACAAGCATATAGGACAAAACAAATATAGGTTTGAATCATGGGA 244  
 QY 3322 GAAGAAGAAATCGGTTTAAATGATTCGTPACCAAGCTTTCTCATGTTGGCCAGTTG 3381  
 Db 243 GAAGAAGAAATCGGTTTAAATGATTCGTPACCAAGCTTTCTCATGTTGGCCAGTTG 184  
 QY 3382 GATGACATAAGAAAACCCCTAGAGTTTGTGCTGATGACACATGACCAAT 3441  
 Db 183 GATGACATAAGAAAACCCCTAGAGTTTGTGCTGATGACACATGACCAAT 124  
 QY 3442 CATAAAGATGCTCAGACAGTGAAGGCTGTCTCAGGAGCTTCTATGAATCCATGTTCCCG 3501  
 Db 123 CATAAAGATGCTCAGACAGTGAAGGCTGTCTCAGGAGCTTCTATGAATCCATGTTCCCG 64  
 QY 3502 ATACCTTCCCAATTTGAATGCGCAAGAGATATCGAAACCGTTTCTTCATATG 3555  
 Db 63 ATACCTTCCCAATTTGAATGCGCAAGAGATATCGAAACCGTTTCTTCATATG 10

RESULT 7  
 BG291336  
 LOCUS  
 DEFINITION 602387205F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4516296 5', mRNA sequence.  
 ACCESSION BG291336  
 VERSION BG291336.1 GI:13049173  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```
|||||
Db 417 CACGACGTAATAATATCCCAACTCAGGAATCCTACTATGATCCCAACCTGNCACC 358
QY 3225 GGTCACTAAAGTCTAGTACAAACTGTAAACAGTAACCTGACAAAATCCCAAGACATA 3284
Db 357 GGTCACTAAAGTCTAGTACAAACTGTAAACAGTAACCTGACAAAATCCCAAGACATA 298
QY 3285 TAAGGACAAAACAAATATAGTTTGAATATCATGGGAGAGAAATGCGCTTTTAAAT 3344
Db 297 TAAGGACAAAACAAATATAGTTTGAATATCATGGGAGAGAAATGCGCTTTTAAAT 238
QY 3345 GATTCTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAAGAAAAACCTAG 3404
Db 237 GATTCTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAAGAAAAACCTAG 178
QY 3405 GAAGTTTGTGGTCTGAATGACAACTATGACCAATATCAATGATGCTCAGACAGTGA 3464
Db 177 GAAGTTTGTGGTCTGAATGACAACTATGACCAATATCAATGATGCTCAGACAGTGA 118
QY 3465 GCCTGTCTCAGGACTTCTATGAATCCATGTTCCTCCATACCTTCCCAATTTGAACTGCC 3524
Db 117 GCCTGTCTCAGGACTTCTATGAATCCATGTTCCTCCATACCTTCCCAATTTGAACTGCC 59
QY 3525 AAGAGATATCGAAACCGTTTCTTCATATGATGATGATGATGATGATGATGATGAT 3571
Db 58 AAGAGATATCGAAACCGTTTCTTCATATGATGATGATGATGATGATGATGATGAT 12

RESULT 9
AW578427
LOCUS RC1-CT0249-120100-022-d01 CT0249 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW578427
ACCESSION AW578427
VERSION AW578427.1 GI:7253476
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t2=RC1-CT0249-
120100-022-d01&t3=2000-01-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 517.
Location/Qualifiers
1. .654
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site: 1; SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 216 a 141 c 127 g 170 t
ORIGIN
```

```
Query Match 15.28; Score 546.2; DB 10; Length 654;
Best Local Similarity 94.1%; Pred. No. 2.1e-117;
Matches 601; Conservative 0; Mismatches 33; Indels 5; Gaps 3;

QY 2716 TACTTCACTAGACAGAATAGAGCCAGATACAAAGAGAGATACATTTGCAGATTCCTC 2775
Db 20 TTCTGGAGTATGACAAAATACCTGGGAGGCAACTAAAGATACATGTCAGATTCCTC 79
QY 2776 AGATATGTAATAAATTTCTAAATAGCAAGTTTGGATTTCACATCGCGGAAGTCCCTGCT 2835
Db 80 AGATATGTAAT - - - AAATCTAAATAGCAAG - TTGGATTTCAGATCGCGGAAGTCCCTGCT 135
QY 2836 CACATGCTCAGATGATTGACCGGATGTTATGCAAGAAGTGCAGAGATATGTTCCCTGAA 2895
Db 136 CACATGCTCAGATGATTGACCGGATGTTATGCAAGAAGTGCAGAGATATGTTCCCTGAA 195
QY 2896 GAATTTGACAAAGAGTGCATTTTACAA - AGTGGCGCATTTGAGGATATGCAAGTTGCGCTT 2954
Db 196 GAATTTGACAAAGAGTGCATTTTACAAAGAGTGCAGGATTCAGGATATGCAAGTTGCGCTT 255
QY 2955 CTCTTATTTTATTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTGA 3014
Db 256 CTCTTATTTTATTTATCTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTGA 315
QY 3015 TGAAGTTGATACAGATCAATCTGCTGCTGACAGAGAAATCCGAACACTGGCTAC 3074
Db 316 TGAAGTTGATACAGACCAATCTGCTGCTGACAGAGAAATCCGAACACTGGCTAC 375
QY 3075 CAGAATTCACGAACCTGCGGTTTAAAGTTTCAGAGTTCGAGGTTGGAACACATGCTAAT 3134
Db 376 CAGAATTCACGAACCTGCGGTTTAAAGTTTCAGAGTTCGAGGTTGGAACACATGCTAAT 435
QY 3135 AAATTTGCTCAAAATGCTTCTGCTGATATCAGCAGCTAAATATATTCACCAACTCA 3194
Db 436 AAATTTGCTCAAAATGCTTCTGCTGATATCAGCAGCTAAATATATTCACCAACTCA 495
QY 3195 GGAATCTCTACTATGATCCCAACCTGCCACGTCACATAAAGTCTAGTAAACACTCTAA 3254
Db 496 GGAATCTCTACTATGATCCCAACCTGCCACGTCACATAAAGTCTAGTAAACACTCTAA 555
QY 3255 ACCAGTAAGTACAAAATTCACAAAGCATATAGGCAACAAACAAATATAGTTTGAAT 3314
Db 556 ACCAGTAAGTACAAAATTCACAAAGCATATCAGCAGCAACAAACAAATATAGTTTGAAT 615
QY 3315 CATGGGAGAGAGAAATGCTGCTTTTAAATGATTCTGATC 3353
Db 616 CATGGGAGAGAGAAATGCTGCTTTATAGGATGATTCTGATC 654

RESULT 10
AW604365
LOCUS RC1-CT0249-290100-022-d01 CT0249 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW604365
ACCESSION AW604365
VERSION AW604365.1 GI:7309106
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 654)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
```

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC1&t3=RC1-CT0249-290100-022-001&t3=2000-01-29&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 6  
High quality sequence stop: 517.  
Location/Qualifiers

## FEATURES

source  
1..654  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0249"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 216 a 141 c 127 g 170 t  
ORIGIN

Query Match 15.2%; Score 546.2; DB 10; Length 654;  
Best Local Similarity 94.1%; Pred. No. 2.1e-117;  
Matches 601; Conservative 0; Mismatches 33; Indels 5; Gaps 3;

```
Qy 2716 TACTTCACTGATGACGAATAGACGACGATACACAGAGATACATTTGCAGATTCCTC 2775
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20 TTCGGAGCTGATGACAAATACTGGGAGGCACTAAAGATACATGTCAGATTCCTC 79

Qy 2776 AGATATGTAAATAAATCTAAATAGCAAGTTTGATTCATCGCGGAAAGTCCCTGCT 2835
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 AGATATGTAAAT---AAATCTAAATAGCAAG-TTGATTCAGATCGCGGAAAGTCCCTGCT 135

Qy 2836 CACATGCTCAGATGATGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCTGAA 2895
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 CACATGCTCAGATGATGACCGGATTTGATGCAAGAACTGCAAGATATGTTCCCTGAA 195

Qy 2896 GAATTTGACAAGACGCTCATTTCAAA-AGTGGCGCATCTTGAGATATGAGTTTGCTT 2954
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 GAATTTGACAAGACGCTCATTTCAAAAGTGGCCATCTTGAGATATGAGTTTGCTT 255

Qy 2955 CTCATTATTTTATTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGA 3014
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 CTCATTATTTTATTCATGAGTGCAGTGCAGCCACTGAATATATCTCAAGTCTTTGA 315

Qy 3015 TGAAGTTGATACAGATCAATCTGCTGTTCTGTGACAGAGAAATCGAACATGGCTAC 3074
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TGAAGTTGATACAGACCAATCTGGTGTCTGTCTGACAGAGAAATCGAACATGGCTAC 375

Qy 3075 CAGAATTCAGAACTGCGGTTAAGTTTGCAGGATTTGACAGGCTGGAAACACATGCTAAT 3134
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 CAGAATTCAGAACTGCGGTTAAGTTTGCAGGATTTGACAGGCTGGAAACACATGCTAAT 435

Qy 3135 AAATGCTCAAAATGCTTCTGCTGATATCAGCAGCTAAATATATTCACCAACTCA 3194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 AAATGCTCAAAATGCTTCTGCTGATATCAGCAGCTAAATATATTCACCAACTCA 495

Qy 3195 GGAATCCTACTATATCCCACTGCCACCGGTCACCTAAAGTCTAGTACAACCTGTAA 3254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 GGAATCCTACTATATCCCACTGCCACCGGTCACCTAAAGTCTAGTACAACCTGTAA 555

Qy 3255 ACCAGTAACAGCAAAATCCACAAGCATATAGGACACAAAATATATAGGTTTGAAT 3314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 ACCAGTAACAGCAAAATCCACAAGCATATACCGGACAAACAAATATAGGTTTGAAT 615

Qy 3315 CATGGGAGAGAAATCGCTTTTAAATGATTCGTAC 3353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 616 CATGGGAGAGAAATCGCTTTTAAATGATTCGTAC 654
```

RESULT 11

## BQ306940

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Matzukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
MEDLINE  
COMMENT  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-BT2002-  
130601-203-b1&t3=2001-06-13&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 574.

Location/Qualifiers  
1..580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT2002"  
/dev\_stage="Adult"

FEATURES  
source  
/note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 194 a 118 c 110 g 158 t  
ORIGIN

Query Match 15.1%; Score 544.2; DB 14; Length 580;  
Best Local Similarity 99.1%; Pred. No. 6.2e-117;  
Matches 568; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 2843 CTCACATGATTCACCGGATTTGTTATGCAAGAACTGCAA-GATATGTTCCCTGAAGAATTT 2901  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 9 CTCACATGATTCACCGGATTTGTTATGCAAGAACTGCAAGGATATGTTCCCTGAAGAATTT 68

Qy 2902 GACACAGCTCATTTACAAAGTGGCCATTTCTGAGGATATGCGATTTGCTCTTAT 2961  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 69 GACAAAACATCTTTTACAAAAGGGCCATTTCTGAGGATATGCGATTTG-CTTCTCTTAT 127

Qy 2962 TTTTATTATCTCATGCTGAGTGCAGCCACTGAATATATCTCAAGTCTTTGATCAAGTT 3021  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 128 TTTTATTATCTCATGCTGAGTGCAGCCACTGAATATATCTCAAGTCTTTGATCAAGTT 187

Qy 3022 GATACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACACTGGCTACCAAGTT 3081  
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 188 GATACAGATCAATCTGGTGTCTTGTCTGACAGAGAAATCCGAACACTGGCTACCAAGTT 247

Qy 3082 CACGAACGCGCTTAAGTTTGCAGGATTTGACAGGCTTGGAAACATGCTAATAAATGC 3141

```

|||||
Db 248 CAGCAACTCCGTTAAGTTTGAGAGTTTGACAGGCTGGACACATGCTATAATAATGC 307
|||||
QY 3142 TCAAAATGCTCTGCTGATATCAGCAGCTAAATAATATTCACCACTCAGGAATCC 3201
|||||
Db 308 TCAAAATGCTCTGCTGATATCAGCAGCTAAATAATATTCACCACTCAGGAATCC 367
|||||
QY 3202 TACTATGATCCCACTCCACCGGTCACATAAAGTCTAGTAACAACTGTAACCCAGTA 3261
|||||
Db 368 TACTATGATCCCACTCCACCGGTCACATAAAGTCTAGTAACAACTGTAACCCAGTA 427
|||||
QY 3262 ACTGACAAATCCACAAAGCATATAAGGACAAACAAATATAGGTTTGAATCATGGGA 3321
|||||
Db 428 ACTGACAAATCCACAAAGCATATAAGGACAAACAAATATAGGTTTGAATCATGGGA 487
|||||
QY 3322 GAAGAGAAATCGCTTTTAAATGATTCGTACCAAGTTTCATGTTGGTGGCCAGTTG 3381
|||||
Db 488 GAAGAGAAATCGCTTTTAAATGATTCGTACCAAGTTTCATGTTGGTGGCCAGTTG 547
|||||
QY 3382 GATGACATAGAGAAACCTAGGAAGTTGTT 3414
|||||
Db 548 GATGACATAGAGAAACCTAGGAAGTTGTT 580

```

```

RESULT 12
AI499228/c 533 bp mRNA linear EST 12-MAY-1999
LOCUS to08f11.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2178477 3'
DEFINITION similar to TR:Q61340 Q61340 KREISLER ;, mRNA sequence.

```

```

ACCESSION AI499228
VERSION AI499228.1 GI:4391210
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
CDNA Library Arrayed by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1973 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 412
POLYA-No.
FEATURES
source Location/Qualifiers
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2178477"
/clone_lib="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DHI08"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; SalI;
Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

```

```

BASE COUNT 116 a 113 c 114 g 188 t 2 others
ORIGIN
Query Match 14.6%; Score 524.6; DB 9; Length 533;
Best Local Similarity 98.9%; Pred. No. 2.5e-112;

```

```

Matches 527; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1978 CCAGAAAGAAACAGCTTCCCGAAGTTTAAAGAGACATGATGTTAACTCAACAAGAGAGACC 2037
|||||
Db 533 CCAGAAAGAAACAGCTTCCCGAAGTTTAAAGAGACATGATGTTAACTCAACAAGNAGAGCN 474
|||||
QY 2038 CAGGAAGAGGTGAAATATCCCTCGTAAATATTTCACTCTCTCCAAAGAGCGCCAGTTG 2097
|||||
Db 473 CAGGAAGAGGTGAAATATCCCTCGTAAATATTTCACTCTCTCCAAAGAGCGCCAGTTG 414
|||||
QY 2098 AGTCTCAATACCTTTGGATTTGCAACTGGACATGGAGACATCACTTTTGAAGGATACAAT 2157
|||||
Db 413 AGTCTCAATACCTTTGGATTTGCAACTGGACATGGAGACATCACTTTTGAAGGATACAAT 354
|||||
QY 2158 TTGTCCAAAGTCAAGCTTGTCTGAGATCATTTCTGATGAACACACAGCATGCTAAATAAAA 2217
|||||
Db 353 TTGTCCAAAGTCAAGCTTGTCTGAGATCATTTCTGATGAACACACAGCATGCTAAATAAAA 294
|||||
QY 2218 AATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAG 2277
|||||
Db 293 AATCAAGCTATAATAACAGATGAAACAAATGACAGTTTGGTGGCTCCACAGGAAAAACAG 234
|||||
QY 2278 GTTCATAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACT 2337
|||||
Db 233 GTTCATAAAAGCATCTTCCCAACAGCTTAGGAGTGTCTGAAAGATTGCAGAGGTTGACT 174
|||||
QY 2338 TTTCTCTGACAGTGAAGTAAAGTGAATGCTCATGACAGGCTCAGATCCACCCCTGGAC 2397
|||||
Db 173 TTTCTCTGACAGTGAAGTAAAGTGAATGCTCATGACAGGCTCAGATCCACCCCTGGAC 114
|||||
QY 2398 TTGGAGACACAGCAAGATTTAGAGTGGAAACTCACACCCCAAAACCATAGCGGAAAT 2457
|||||
Db 113 TTGGAGACACAGCAAGATTTAGAGTGGAAACTCACACCCCAAAACCATAGCGGAAAT 54
|||||
QY 2458 GTGCAAAAGAAAGCCCACTCTCTGATTTTCCACTGGAAAGCCAGATGAC 2510
|||||
Db 53 GTGCAAAAGAAAGCCCACTCTCTGATTTTCCACTGGAAAGCCAGATGAC 1
|||||
RESULT 13
BE736417 572 bp mRNA linear EST 15-SEP-2000
LOCUS 601306591F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3640726 5',
DEFINITION mRNA sequence.
ACCESSION BE736417
VERSION BE736417.1 GI:10150409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10C345 row: f column: 23
High quality sequence stop: 572.
Location/Qualifiers
1..572
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3640726"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
FEATURES
source
1..572

```

```

/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT      197 a 124 c 114 g 136 t 1 others
ORIGIN
Query Match      14.3%; Score 514.4; DB 12; Length 572;
Best Local Similarity 98.2%; Pred. No. 6.1e-110;
Matches 541; Conservative 0; Mismatches 7; Indels 3; Gaps 2;
QY 3048 TGACAGAGAAATCCGAACTGCTACAGAAATTCAGAACTCCCGTTAAGTTTCAGGA 3107
Db 1 TGACAGAGAAATCCGAACTGCTACAGAAATTCAGAACTCCCGTTAAGTT--GCAGG 58
QY 3108 TTTGACAGGTTCTGGAACACATGCTAATAAATGCTCAAAATGCTTCCTGCTGATATCAC 3167
Db 59 ATTGACAGGTTCTGGAACACATGCTAATAA--TGCTCAAAATGCTTCCTGCTGATATCAC 117
QY 3168 GCAGCTAAATAATATTCACCACTCAGAAATCTCTACTATGATGCCAAGTCCACCGGT 3227
Db 118 GCAGCTAAATAATATTCACCACTCAGAAATCTCTACTATGATGCCAAGTCCACCGGT 177
QY 3228 CACTAAAAGTCTAGTAACAACTGTAAACAGTAACTGACAAATCCACAAAGCATATAA 3287
Db 178 CACTAAAAGTCTAGTAACAACTGTAAACAGTAACTGACAAATCCACAAAGCATATAA 237
QY 3288 GGACAAAACAATATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAATGAT 3347
Db 238 GGACAAAACAATATAGTTTGAATCATGGGAGAGAAAGAAATCGCTTTTAAATGAT 297
QY 3348 TCGTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAGAAACCCCTAGGAA 3407
Db 298 TCGTACCAAGTTTCTCATGTGGTGGCCAGTTGGATGACATAGAAAGAAACCCCTAGGAA 357
QY 3408 GTTTGTTGGCTGAATGACAACTGACCAATCATAAAGATGCTCAGACAGTGAAGGC 3467
Db 358 GTTTGTTGGCTGAATGACAACTGACCAATCATAAAGATGCTCAGACAGTGAAGGC 417
QY 3468 TGTCTCAGGACTCTATGATTCATGTTCCCATACCTTCCCAATTTGAAGTCCCAAG 3527
Db 418 TGTCTCAGGACTCTATGATTCATGTTCCCATACCTTCCCAATTTGAAGTCCCAAG 477
QY 3528 AGAGTATCGAAACCGTTTCTTCATATGATGAGCTGAGGAATGGAGGCTTATCGAGA 3587
Db 478 AGAGTATCGAAACCGTTTCTTCATATGATGAGCTGAGGAATGGAGGCTTATCGAGA 537
QY 3588 CAAATGGAAGT 3598
Db 538 CAAATGGAAGT 548

```

```

RESULT 14
BF739893/c
LOCUS
DEFINITION
709 bp mRNA linear EST 10-JAN-2001
similar to TR:Q9ULL2 Q9ULL2 KIAA1208 PROTEIN ; mRNA sequence.
ACCESSION BF739893
VERSION BF739893.1 GI:12066557
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 709)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

## JOURNAL COMMENT

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov

Seq primer: -40UP from Gibco  
 High quality sequence stop: 460.

## FEATURES source

1..709  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3576619"  
 /clone\_lib="NCI\_CGAP\_Kid11"  
 /lab\_host="DH10B"  
 /notes="Organ: kidney; Vector: pOTB73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 174 a 141 c 135 g 259 t  
 ORIGIN

Query Match 14.1%; Score 509; DB 12; Length 709;  
 Best Local Similarity 98.1%; Pred. No. 1.1e-108;  
 Matches 515; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
 QY 2164 AAGTCAGCCTTGCTGAGATCATTTCTGATGAACCTCAGCAGCATGCTAAATATAAAATCAA 2223  
 Db 709 AAGTCAGCCTTGCTGAGATCATTTCTGATGAACCTCAGCAGCATGCTAAATATAAAATCAA 650  
 QY 2224 GCTATATAACAGATGAACAAATAGCAGTTTGGTGGCTCCACAGGAAACAGGTTTCAAT 2283  
 Db 649 CTTATATAACAGATGAACAAATAGCAGTTTGGTGGCTCCACAGGAAACAGGTTTCAAT 590  
 QY 2284 AAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGCAGAGTTGACTTTTCCT 2343  
 Db 589 AAAAGCATCTTGCCAAACAGCTTAGGAGTGTCTGAAAGATTGCAGAGTTGACTTTTCCT 530  
 QY 2344 GCAGTGTGTAAGTGAATGATGTCATGACCGGGTCCAGAAATCCACCCCTGGACTTTGGAG 2403  
 Db 529 GCAGTGTGTAAGTGAATGATGTCATGACCGGGTCCAGAAATCCACCCCTGGACTTTGGAG 470  
 QY 2404 ACCACACAGATTTAGAGTGGAACTCAGCCCAACCAACCATAGCGGAAATGTGACA 2463  
 Db 469 ACCACACAGATTTAGAGTGGAACTCAGCCCAACCAACCATAGCGGAAATGTGACA 410  
 QY 2464 AAAAGAAAGCCCTCATCTCTGATTGTTCCACTGGAAGCCAGATGACAAAGAAAGAAA 2523  
 Db 409 AAAGAAAGCCCTCATCTCTGATTGTTCCACTGGAAGCCAGATGACAAAGAAAGAAA 350  
 QY 2524 ATCAGAGGGAAGAAAGAAAGACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGC 2583  
 Db 349 ATCAGAGGGAAGAAAGAAAGACAGTAGAATGGAGGAAATGCTGAAAATCACATAGGC 290  
 QY 2584 GTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATTCACAGATAGTTACTTGGGCTTT 2643  
 Db 289 GTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATTCACAGATAGTTACTTGGGCTTT 230  
 QY 2644 TTGCCATGGGAGAAAGAAAGATTTTCCCTAGATCTTCTCGACGAA 2688  
 Db 229 TTGCCATGGGAGAAAGAAAGATTTTCCCAAGATCTTCTCGACGTA 185



## RESULT 15

LOCUS BG085937 673 bp mRNA linear EST 26-JAN-2001  
 DEFINITION H3119H05-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone

ACCESSION BG085937

VERSION H3119H05 5', mRNA sequence.

KEYWORDS EST.

SOURCE BG085937.1 GI:12568501

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 673)

Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka

T.S., Carter, M.G. and Ko, M.S.H.

Verification and initial annotation of NIA mouse 15K cDNA clone set

Unpublished (2001)

Other ESTs: H3119H05-3

Contact: George J. Kargul

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please

visit <http://igsun.grc.nia.nih.gov/cDNA/15K.html> for details.

Plate: H3119 row: H column: 05

Seq primer: -21M13 Reverse

High quality sequence stop: 673

POLYA=No.

FEATURES

Location/Qualifiers

1..673

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="niaEST:H3119H05-5"

/db\_xref="taxon:10090"

/clone="H3119H05"

/clone.lib="NIA Mouse 15K cDNA Clone Set"

/sex="Clones arrayed from a variety of cDNA libraries"

/dev\_stage="Clones arrayed from a variety of cDNA

libraries"

/lab\_host="DH10B"

/note="Vector: pSPORT1; Site1: SalI; Site2: NotI; This

clone is among a rearranged set of 15,247 clones from 11

embryo cDNA libraries (including preimplantation stage

embryos from unfertilized egg to blastocyst, embryonic

part of E7.5 embryos, extraembryonic part of E7.5 embryos

, and E12.5 female mesonephros/gonad) and one newborn

ovary cDNA library. Average insert size 1.5 kb. All

source libraries are cloned unidirectionally with Oligo(dT

) -Not primers. References include: (1) Genome-wide

expression profiling of mid-gestation placenta and embryo

using a 15,000 mouse developmental cDNA microarray, 2000,

Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)

Large-scale cDNA analysis reveals phased gene expression

patterns during preimplantation mouse development, 2000,

Development, 127: 1737-1749; (3) Genome-wide mapping of

unselected transcripts from extraembryonic tissue of

7.5-day mouse embryos reveals enrichment in the t-complex

and under-representation on the X chromosome, 1998, Hum

Mol Genet 7: 1967-1978."

Mol Genet 7: 1967-1978."

BASE COUNT 207 a 167 c 136 g 163 t

ORIGIN

Query Match

Best Local Similarity 14.1%; Score 507.2; DB 12; Length 673;

Matches 551; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 2975 TCAGTGCAGTCGACGACCAATATCTCAAGTCTTTGATGAGTTCATACAGATCAAT 3034

Db 1 TCAGTGCAGTCGACGACCAATATCTCAAGTCTTTGATGAGTTCATACAGATCAAT 60

QY 3035 CTGGTGTCTTGTCTGACAGAGAAATCCGAAACACTGGCTACCAGAAATTCACGAACTGCCGT 3094

Db 61 CTGGTGTCTTGTCTGATAGGGAATTCGAAACACTGGCCAGAGAAATTCACGACCTACCTT 120  
 QY 3095 TAAGTTTGCAGGATTTGACAGGCTCGGAACACATGCTTAATAAATGCTCAAAAATGCTTC 3154  
 Db 121 TAAGCTTGCAGGATTTGACAGGCTTTGGAACACATGTTTAATAAATGCTCAAAAATGCTTC 180  
 QY 3155 CTGCTGATATCACGCAGCTAAATAATATCCACCAACTCAGGAATCCTTACTATGATCCCA 3214  
 Db 181 CCGTATATCCTCAACTCAACCAACTCCCGGACTCAGGAAGCATACTACGCCCA 240  
 QY 3215 ACCTGCCACCGGTCTCACTAAAGTCTAGTAAACAACTGTAAACCACTGTAACCTGACAAAAATCC 3274  
 Db 241 ACCTGCCCTCCGCTCACTAAAGTCTTGTCAACCACTGTAAGCCAGTAAGTACAGATCC 300  
 QY 3275 ACAAGCATATAAGGAGCAAAACAAATATAGTTTGAATCATGCGGAGAGAGAAATCG 3334  
 Db 301 ACAAGGCTATATAAGAGCAAAACAAATACAGGTTTGAATCATGCGGAGAGAGAAATCG 360  
 QY 3335 CTTTTAAATGATTTCGTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAA 3394  
 Db 361 CTTTCAAGATATACGAACCAATGTTTCTCATGTGGTGGCCAGTTGGATGACATCAGAA 420  
 QY 3395 AAAACCCCTAGGAAGTCTTGTTCCTGAATGACAACTGACCACTGACCAATCATAAAGATGCTC 3454  
 Db 421 AAAACCCCGAGGAGTCTGTTGTCTGAATGACAACTGACCACTGACCAATCATAAAGATGCTC 480  
 QY 3455 AGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAAT 3514  
 Db 481 GGACAGTGAAGGCTGTTCTCAGGAGCTTCTATGATCCATGTTTCCCATACCTTCCCAAT 540  
 QY 3515 TTGAACGCCAAGAGAGTATCGAAACGTTTCTTCATATGTCATGACCTGCAGGAATGGA 3574  
 Db 541 TTGAGCTGCCAAGAGAGTATCGAAACGCTTCTGACATGTCATGACCTGCAGGAATGGA 600  
 QY 3575 GGGCTTATCGAGACAAATTTGAAAT 3598  
 Db 601 GGGCATATCGAGACAACTGAAAT 624

Search completed: April 13, 2003, 01:34:36  
 Job time : 3215 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 12, 2003, 22:47:35 ; Search time 6160 Seconds  
(without alignments)  
17008.133 Million cell updates/sec

Title: US-10-023-888-1  
Perfect score: 3600  
Sequence: 1 atggagacagacacactct.....atcgagacaaattgaagtag 3600

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.pat.\*
- 6: gb.ph.\*
- 7: gb.pl.\*
- 8: gb.pr.\*
- 9: gb.ro.\*
- 10: gb.sts.\*
- 11: gb.sy.\*
- 12: gb.un.\*
- 13: gb.vi.\*
- 14: gb.ba.\*
- 15: em.fun.\*
- 16: em.in.\*
- 17: em.mu.\*
- 18: em.ov.\*
- 19: em.pat.\*
- 20: em.ph.\*
- 21: em.pl.\*
- 22: em.ro.\*
- 23: em.sts.\*
- 24: em.un.\*
- 25: em.vi.\*
- 26: em.htg\_hum.\*
- 27: em.htg\_inv.\*
- 28: em.htg\_other.\*
- 29: em.htg\_mus.\*
- 30: em.htg\_pln.\*
- 31: em.htg\_rod.\*
- 32: em.htg\_mam.\*
- 33: em.htg\_vrt.\*
- 34: em.sy.\*
- 35: em.htgo\_hum.\*
- 36: em.htgo\_mus.\*
- 37: em.htgo\_pln.\*
- 38: em.htgo\_rod.\*
- 39: em.htgo\_mam.\*
- 40: em.htgo\_vrt.\*
- 41: em.htgo\_sy.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3472	96.4	4369	6	AX468102	Sequence
2	3100.4	86.1	4277	6	AX468106	Sequence
3	2685.4	74.6	4511	9	AB033034	Homo sapi
4	2404.8	66.8	2603	9	AK056137	Homo sapi
5	1103.6	30.7	162427	9	AC010205	Homo sapi
6	1103.6	30.7	174231	9	AC063950	Homo sapi
c	1103.2	30.6	177364	9	AC005409	Homo sapi
c	849.8	23.6	1846	10	MUSBDL2	Mus Musculu
8	592.8	16.5	2428	9	HS0802690	Homo sapi
c	499	13.9	141215	2	AC121402	Rattus no
10	462.4	12.8	281892	2	AC125486	Mus muscu
11	403.8	11.2	2332	9	HS0803479	Homo sapi
12	266	7.4	2076	9	AK001821	Homo sapi
13	214.4	6.0	142667	10	AL135758	Mouse DNA
14	184.8	5.1	12996	10	AC005501	Mus muscu
c	183.4	5.1	198967	2	AC124926	Rattus no
16	165.6	4.6	281892	2	AC125486	Mus muscu
c	157.6	4.4	141215	2	AC121402	Rattus no
c	157.6	4.4	177607	2	AC110966	Rattus no
19	157.6	4.4	177607	2	AC110966	Rattus no
20	138	3.8	2610	3	AY089618	Drosophil
21	138	3.8	51430	2	AC017981	Drosophil
22	138	3.8	189757	3	AC008186	Drosophil
c	138	3.8	243313	3	AE003834	Drosophil
c	119.4	3.3	46555	2	AC100621	Mus muscu
26	102.6	2.8	42816	1	SCC88	Streptomy
c	100.4	2.8	36849	1	SC1C3	Streptomy
27	91.4	2.5	15347	6	AX283713	Sequence
28	91.4	2.5	28170	6	AX283721	Sequence
29	76.4	2.1	14624	1	SSA17900	Streptococ
30	76.4	2.1	20555	6	AX283717	Sequence
31	74.4	2.1	18274	6	AX283720	Sequence
32	73	2.0	340806	1	NNA12491	Sequence
33	72	2.0	780	9	BC002779	Homo sapi
34	71.4	2.0	5064	1	AF019760	Neisseria
35	69.6	1.9	4798	1	AF518558	Actinobac
c	67	1.9	7218	6	I66494	Sequence 14
37	66.6	1.8	5070	6	AX234391	Sequence
38	65.6	1.8	1779	6	AX451462	Sequence
39	65.6	1.8	2670	6	E59779	Soluble neu
40	65.2	1.8	624	6	AX084406	Sequence
41	65.2	1.8	624	6	AX128409	Sequence
42	65.2	1.8	5732	1	AF112478	Neisseria
43	64	1.8	5732	6	AX027904	Sequence
44	64	1.8	5732	12	AX027910	Sequence
45	64	1.8	5732	12	AX027910	Sequence

ALIGNMENTS

RESULT 1  
AX468102  
LOCUS AX468102  
DEFINITION Sequence 25 from Patent WO0250279.  
ACCESSION AX468102  
VERSION AX468102.1  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1  
Baughin, M.R., Lu, Y., Arvizu, C., Ramkumar, J., Yao, M.G.,  
Policky, J.L., Walia, N.K., Tribouley, K.M., Yue, H., Batra, S.,  
Ding, L., Lal, P.G., Borowsky, M.L., Lu, D.A., Gandhi, A.R.,

Griffin, J.A., Xu, Y., Azimzai, Y., Gietzen, K.J., Tang, Y.T.,  
Warren, B.A., Mason, P.M., Burford, N.M., Hafalia, A.J., Lee, E.A.,  
Yang, J., Gorvad, A.E., Emerling, B.M., Marquis, J.P., Lee, S.Y.,  
Swarnakar, A. and Reddy, R.  
Nucleic acid-associated proteins  
Patent: WO 0250279-A 25 27-JUN-2002;  
Incyte Genomics, Inc. (US)  
Location/Qualifiers  
source  
1. 4369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2749402CB1"

BASE COUNT 1339 a 912 c 978 g 1140 t  
ORIGIN

Query Match 96.4%; Score 3472; DB 6; Length 4369;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 3481; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 103 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAAGTCC 162  
|||||  
Db 266 AGCCGAGATCAATACCATGTTTGTGATTCCTATAGAGACAATATGCTGGAAAGTCC 325  
|||||  
QY 163 TTTCAGATCGCTTGTGTCGCCATGCCGATTGACGTGTTTACACCTGGGTGAATGCC 222  
|||||  
Db 326 TTTCAGATCGCTTGTGTCGCCATGCCGATTGACGTGTTTACACCTGGGTGAATGCC 385  
|||||  
QY 223 ACAGATCTTGAACACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGAGAGAGCAG 282  
|||||  
Db 386 ACAGATCTTGAACACTACTGAAGAACTACAGCAGGTTCAGAGAACAGATGAGAGAGCAG 445  
|||||  
QY 283 AAAGCAATGAGAGAAATCCTTGGGAAAAACACACGGAACCTACTAAGAGAGTGAAG 342  
|||||  
Db 446 AAAGCAATGAGAGAAATCCTTGGGAAAAACACACGGAACCTACTAAGAGAGTGAAG 505  
|||||  
QY 343 CAGTTAGAGTCTTCTTAACACACTGCAATTAAGTGCCCAATGCTTGTCCCTGGACCCAGCC 402  
|||||  
Db 506 CAGTTAGAGTCTTCTTAACACACTGCAATTAAGTGCCCAATGCTTGTCCCTGGACCCAGCC 565  
|||||  
QY 403 CTGCCAGCAACATCACCTCTGAAGGACCTGCCATCTCTTTATCCTCTTTTCAATCTGCC 462  
|||||  
Db 566 CTGCCAGCAACATCACCTCTGAAGGACCTGCCATCTCTTTATCCTCTTTTCAATCTGCC 625  
|||||  
QY 463 AGTGACATTTCAATGTTGGAACCAACCAAAACCCCTCTACCAATGCTCAGTTGTTGT 522  
|||||  
Db 626 AGTGACATTTCAATGTTGGAACCAACCAAAACCCCTCTACCAATGCTCAGTTGTTGT 685  
|||||  
QY 523 TTTCAGTACTTAAGGATGTTGAAGATGCCACCTGCGACTGCTTAAAGGAAATAGCAGA 582  
|||||  
Db 686 TTTCAGTACTTAAGGATGTTGAAGATGCCACCTGCGACTGCTTAAAGGAAATAGCAGA 745  
|||||  
QY 583 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAGTCTAATG 642  
|||||  
Db 746 CAGACAGTATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAGTCTAATG 805  
|||||  
QY 643 CAAGATTTGGCTTCTCCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAA 702  
|||||  
Db 806 CAAGATTTGGCTTCTCCTGAGTGGATTTCCACCAACATTCAGGAAACAAATCAACTAAA 865  
|||||  
QY 703 ACAGATTTGCCAGAAAATCTTCTCTTAAAGTCAAACTGTTGCACTGCTATTCAGAGGCC 762  
|||||  
Db 866 ACAGATTTGCCAGAAAATCTTCTCTTAAAGTCAAACTGTTGCACTGCTATTCAGAGGCC 925  
|||||  
QY 763 AGTGTAGGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 822  
|||||  
Db 926 AGTGTAGGCTTCTAAAACCTGAATAACCCCAAGGATTTTCAAGAAATGAATAAGCAAACT 985  
|||||  
QY 823 AAGAAGACATGACATTTGATGGAAGAACTGACCAATGCTGATATTTATATGG 882  
|||||  
Db 986 AAGAAGACATGACATTTGATGGAAGAACTGACCAATGCTGATATTTATATGG 1045  
|||||  
QY 883 GATCTGAGCCCATCAGCCACTTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTAA 942  
|||||

Db 1046 GATCTGAGGCCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTCTGTTTGA 1105  
QY 943 GATAACGAAGAACTGAGGTACTCATTCGCGATCTATCGAGAGCATGACCATGGTTCGG 1002  
|||||  
Db 1106 GATAACGAAGAACTGAGGTACTCATTCGCGATCTATCGAGAGCATGACCATGGTTCGG 1165  
|||||  
QY 1003 AATATTTTTCATGTCACCAACGGGAGATTCATCCTGCTGAACTTGACAATCCTGA 1062  
|||||  
Db 1166 AATATTTTTCATGTCACCAACGGGAGATTCATCCTGCTGAACTTGACAATCCTGA 1225  
|||||  
QY 1063 GTGACAAATAGTACACACAGGATGTTTTCGAAAATTTTCAGCCACTTGCTACCTTATG 1122  
|||||  
Db 1226 GTGACAAATAGTACACACAGGATGTTTTCGAAAATTTTCAGCCACTTGCTACCTTATG 1285  
|||||  
QY 1123 TCACCTGCTATTGAAAAGTCACTCATCGCATCGAAGGCTGTCCACAGAAGTTTATTTAC 1182  
|||||  
Db 1286 TCACCTGCTATTGAAAAGTCACTCATCGCATCGAAGGCTGTCCACAGAAGTTTATTTAC 1345  
|||||  
QY 1183 CTAAATGATGATGTCATGTTTGGGAAGATGTCGTGCCAGATGATTTTACAGTCACTCC 1242  
|||||  
Db 1346 CTAAATGATGATGTCATGTTTGGGAAGATGTCGTGCCAGATGATTTTACAGTCACTCC 1405  
|||||  
QY 1243 AAAGGCCAGAAGGTTTATTTGACATGGCCTGTGCCAAACTGTGCCGAGGCTGCCAGT 1302  
|||||  
Db 1406 AAAGGCCAGAAGGTTTATTTGACATGGCCTGTGCCAAACTGTGCCGAGGCTGCCAGT 1465  
|||||  
QY 1303 TCTTGGATTAAAGGATGGCTATTGTGACAAAGGCTTGTAAATTCAGCCTGCGATTGGAT 1362  
|||||  
Db 1466 TCTTGGATTAAAGGATGGCTATTGTGACAAAGGCTTGTAAATTCAGCCTGCGATTGGAT 1525  
|||||  
QY 1363 GGTGGGATTTGCTCTGGAAACAGTGGAGGAGTTCGTATATTTGAGGAGGTGGAGTACT 1422  
|||||  
Db 1526 GGTGGGATTTGCTCTGGAAACAGTGGAGGAGTTCGTATATTTCCAGGAGGTGGAGTACT 1585  
|||||  
QY 1423 GGAGTATTTGGAGTGGACAGCCCTGGCAGTTTGTGGAGGAATAAACAGTCTCTTAC 1482  
|||||  
Db 1586 GGAGTATTTGGAGTGGACAGCCCTGGCAGTTTGTGGAGGAATAAACAGTCTCTTAC 1645  
|||||  
QY 1483 TGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1542  
|||||  
Db 1646 TGTAATCAGGATGTCGGAATTCCTGGCTCGCTGATTAAGTTCTGTGACCAAGCATGCAAT 1705  
|||||  
QY 1543 GTCTTGTCTGTGGGTTTGATGCTGGCAGTGTGGCAAGATCATTTTCATGAATGTAT 1602  
|||||  
Db 1706 GTCTTGTCTGTGGGTTTGATGCTGGCAGTGTGGCAAGATCATTTTCATGAATGTAT 1765  
|||||  
QY 1603 AAAGTGAATCTCTCCCAACAGACTCACTATATTTATTTCCAAAAGGTGAATGCGCTGCT 1662  
|||||  
Db 1766 AAAGTGAATCTCTCCCAACAGACTCACTATATTTATTTCCAAAAGGTGAATGCGCTGCT 1825  
|||||  
QY 1663 TATTTTCAGCTTTGCAAGATAGCCAAAGAGGAGTTGAAGTGCCTATAGTGAACATCCA 1722  
|||||  
Db 1826 TATTTTCAGCTTTGCAAGATAGCCAAAGAGGAGTTGAAGTGCCTATAGTGAACATCCA 1885  
|||||  
QY 1723 ATAAATTCGACATGCTTCTATTGCCACAGTGGAAACCATCCACTCATATATGCACAGT 1782  
|||||  
Db 1886 ATAAATTCGACATGCTTCTATTGCCACAGTGGAAACCATCCACTCATATATGCACAGT 1945  
|||||  
QY 1783 GGAATGAATGCCACCAATACATATTTTAACTCTCACGTTTCAAAAATACAAACCATGAAG 1842  
|||||  
Db 1946 GGAATGAATGCCACCAATACATATTTTAACTCTCACGTTTCAAAAATACAAACCATGAAG 2005  
|||||  
QY 1843 TTCAAAATCGAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTACG 1902  
|||||  
Db 2006 TTCAAAATCGAGATAACAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTACG 2065  
|||||  
QY 1903 GCCCAGAGGTTTACGAAATTTAGTTAGTCCATTAACACTTCTTCCAGAGCGGAATC 1962  
|||||  
Db 2066 GCCCAGAGGTTTACGAAATTTAGTTAGTCCATTAACACTTCTTCCAGAGCGGAATC 2125  
|||||  
QY 1963 CTTTTCAGGATATTCCTCCAAAGAAAACCTTCCCGAAGTTTAAAGACATGATGTTTAC 2022  
|||||  
Db 2126 CTTTTCAGGATATTCCTCCAAAGAAAACCTTCCCGAAGTTTAAAGACATGATGTTTAC 2185  
|||||

QY 2023 TCACACAGGAGAGCCAGGAGAGGTCACAAATTCCTCGTAAATATTTTCACTCCTTCCA 2082  
Db 2186 TCACACAGGAGAGAGCCAGGAGAGGTCACAAATTCCTCGTAAATATTTTCACTCCTTCCA 2245  
QY 2083 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2142  
Db 2246 AAAGACGCCAGTTGAGTCTCAATACCTTGGATTTGCAACTGGAACATGGAGACATCACT 2305  
QY 2143 TTGAAGGATACAAATTTGTCACAGTCAGCTTCCCTGAGATCAATTTCTGATGACTCACAG 2202  
Db 2306 TTGAAGGATACAAATTTGTCACAGTCAGCTTCCCTGAGATCAATTTCTGATGACTCACAG 2365  
QY 2203 CATGCTAAATAAATAAATCAAGCTATAATAACAGATGAACAAATCACAGTTTGGTGGCT 2262  
Db 2366 CATGCTAAATAAATAAATCAAGCTATAATAACAGATGAACAAATCACAGTTTGGTGGCT 2425  
QY 2263 CCACAGGAAACAGAGTTTCATAAAGCATCTTCCCAACAGCTTAGAGTGTCTGAAAGA 2322  
Db 2426 CCACAGGAAACAGAGTTTCATAAAGCATCTTCCCAACAGCTTAGAGTGTCTGAAAGA 2485  
QY 2323 TTGCAGAGTTGACTTTTCTGTCAGTGAAGTGTAAAAGTGAATGGTCAATGACCAAGGTGAC 2382  
Db 2486 TTGCAGAGTTGACTTTTCTGTCAGTGAAGTGTAAAAGTGAATGGTCAATGACCAAGGTGAC 2545  
QY 2383 AATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGAACACTCACACCCCAAAA 2442  
Db 2546 AATCCACCCCTGGACTTGGAGACACAGCAAGATTTAGAGTGAACACTCACACCCCAAAA 2605  
QY 2443 ACCATAGGCGGAATGTGACAAAAGAAAGCCGCCCTCTCTGATTTTCCACTGGAAGC 2502  
Db 2606 ACCATAGGCGGAATGTGACAAAAGAAAGCCGCCCTCTCTGATTTTCCACTGGAAGC 2665  
QY 2503 CAGATGACAAAAGAAAGAAATCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2562  
Db 2666 CAGATGACAAAAGAAAGAAATCACAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2725  
QY 2563 AATGCTGAAATCACATAGGCGTTACTGAGTGTACTTGGAGAAAGCTGCAGCATTTAC 2622  
Db 2726 AATGCTGAAATCACATAGGCGTTACTGAGTGTACTTGGAGAAAGCTGCAGCATTTAC 2785  
QY 2623 ACAGATAGTTACTTGGCGTTTGGCCATGGGAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2682  
Db 2786 ACAGATAGTTACTTGGCGTTTGGCCATGGGAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2845  
QY 2683 GACCAAGAGAGTCATGACACACAAATGGCCCTACTTCTCATGATGACGAAGATAGAGC 2742  
Db 2846 GACCAAGAGAGTCATGACACACAAATGGCCCTACTTCTCATGATGACGAAGATAGAGC 2905  
QY 2743 AGATACAAAGAGATACATTTGCAGATTTCCCTCAGATATGTAAATAAATTTCTAAATAGC 2802  
Db 2906 AGGCAACTAAAGATACATTTGCAGATTTCCCTCAGATATGTAAATAAATTTCTAAATAGC 2965  
QY 2803 AAGTTTGGATTCATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGCTCAGATGCTCAGAT 2862  
Db 2966 AAGTTTGGATTCATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGCTCAGATGCTCAGAT 3025  
QY 2863 GTTATGCAAGAACTGCAGATATGTTCCCTGAGAAATTTGACAAAGAGCTCATTTACAAA 2922  
Db 3026 GTTATGCAAGAACTGCAGATATGTTCCCTGAGAAATTTGACAAAGAGCTCATTTACAAA 3085  
QY 2923 GTGGCCCATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 2982  
Db 3086 GTGGCCCATTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAGTGA 3145  
QY 2983 GTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGT 3042  
Db 3146 GTGCAGCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGT 3205  
QY 3043 TTGTCTGACAGAAATCCGAAACACTGGGTACAGAAATTCAGAAATTCAGAAATTCAGT 3102  
Db 3206 TTGTCTGACAGAAATCCGAAACACTGGGTACAGAAATTCAGAAATTCAGAAATTCAGT 3265

QY 3103 CAGGATTTCAGAGTCTGAAACACATGCTAATAAATTCCTCAAAAATTCCTTCCTGCTGAT 3162  
Db 3266 CAGGATTTCAGAGTCTGAAACACATGCTAATAAATTCCTCAAAAATTCCTTCCTGCTGAT 3325  
QY 3163 ATCCAGCAGCTAAATAATATTTCCCAACTCAGGAATCCTACTATGATCCCAACTGCA 3222  
Db 3326 ATCCAGCAGCTAAATAATATTTCCCAACTCAGGAATCCTACTATGATCCCAACTGCA 3385  
QY 3223 CCGGTCTACTAAAGTCTAGTAACAACTGTAAACAGTAACAGTAACAGTAACAGTAACAG 3282  
Db 3386 CCGGTCTACTAAAGTCTAGTAACAACTGTAAACAGTAACAGTAACAGTAACAGTAACAG 3445  
QY 3283 TATAAGGACAAAACAAATATAGTTTCAAAATCATGGAGAGAAAGAAATCGCTTTTAAA 3342  
Db 3446 TATAAGGACAAAACAAATATAGTTTCAAAATCATGGAGAGAAAGAAATCGCTTTTAAA 3505  
QY 3343 ATGATTCTGACCAAGCTTCTCATGTGTTGCCAGTTGGATGACATGAAGAAAAACCT 3402  
Db 3506 ATGATTCTGACCAAGCTTCTCATGTGTTGCCAGTTGGATGACATGAAGAAAAACCT 3565  
QY 3403 AGGAAGTTTGTGCTGAATGACAACTTGAACCAATCAATGAAGTCTCAGACAGTG 3462  
Db 3566 AGGAAGTTTGTGCTGAATGACAACTTGAACCAATCAATGAAGTCTCAGACAGTG 3625  
QY 3463 AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGTG 3522  
Db 3626 AAGGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAAGTG 3685  
QY 3523 CCAAGAGAGTATCGAAACCGTTTCCCTCATATGATGATGATGATGATGATGATGATGAT 3582  
Db 3686 CCAAGAGAGTATCGAAACCGTTTCCCTCATATGATGATGATGATGATGATGATGATGAT 3745  
QY 3583 CGAGACAAATTTGAAGT 3598  
Db 3746 CGAGACAAATTTGAAGT 3761

## RESULT 2

LOCUS AX468106 4277 bp DNA linear PAT 16-JUL-2002  
DEFINITION Sequence 29 from Patent WO0250279.  
ACCESSION AX468106  
VERSION AX468106.1 GI:21900980  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Baughin, M.R., Lu, Y., Arvizu, C., Ramkumar, J., Yao, M.G.,  
Policky, J.L., Walla, N.K., Tribouley, K.M., Yue, H., Batra, S.,  
Ding, L., Lal, P.G., Borowsky, M.L., Lu, D.A., Gandhi, A.R.,  
Griffin, J.A., Xu, Y., Azimzai, Y., Gietzen, K.J., Tang, Y.T.,  
Warren, B.A., Mason, P.M., Burford, N., Hafalia, A.J., Lee, E.A.,  
Yang, J., Gorvad, A.E., Emerling, B.M., Marquis, J.P., Lee, S.Y.,  
Swarnakar, A. and Reddy, R.  
TITLE Nucleic acid-associated proteins  
JOURNAL Patent: WO 0250279-A 29 27-JUN-2002;  
Incyte Genomics, Inc. (US)  
FEATURES  
Location/Qualifiers  
source  
1. 4277  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7503560CB1"  
BASE COUNT 1325 a 891 c 925 g 1135 t 1 others  
ORIGIN  
Query Match 86.1%; Score 3100.4; DB 6; Length 4277;  
Best Local Similarity 94.4%; Pred. No. 0;  
Matches 3300; Conservative 0; Mismatches 16; Indels 180; Gaps 1;  
QY 103 AGCCGAGATCAATACCATGTTTGTGTTTCTCATGACATATGCTGGAAGTCC 162  
|||||





RRFVETHQKTIIGNVTKERPPSLIVPLESOMTKKKITCKEKENSRMEENAEHIGV  
TEVLLGKRLHYDTSYLGFLPWEKKYFQDLDEESLKTQLAYFDSKNTGRLQKXD  
FADSLRVNKLINSKFQTSRKVPAPHPHMDIRVMOELQDMPEFEDKTSRPHVRHS  
EDMOPAFSYLYLMSAVOPLNIQOVDEVDTDOSGLVSOREIETLATHLEPLSLQD  
LTGLEHLINCSKMLPADITQNLNIPPTQESYDPLNPVTKSLVTRNCRPVDKIHKA  
YKDKNTRFEIMGEETAFKMRITNVSHVVGQDDIRKNRPKEVCLNDINIDHNHDKAQ  
TVKAVLDRFYESMFPPIPSQEPREYRNRELMHELOEWRAYRDKLKFTHCVLATLI  
MFITFSFABQLIALKRKKIPRRRIKEASPNRIRV"  
BASE COUNT 1382 a 898 c 938 g 1293 t  
ORIGIN

Query Match 74.68; Score 2685.4; DB 9; Length 4511;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 2695; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY	888	GAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTCCAGTCGTTTGAAGATAA	947
DB	1	GAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTCCAGTCGTTTGAAGATAA	60
QY	948	CGAAGACTGAGGTACTCATTCGATCTATCGAGAGCATGCACCATGGGTTCGGAATAT	1007
DB	61	CGAAGACTGAGGTACTCATTCGATCTATCGAGAGCATGCACCATGGGTTCGGAATAT	120
QY	1008	TTTCATTCACCAACGGCAGATTCATCTGCTGAACCTTGACAACTCCTCGAGTGAC	1067
DB	121	TTTCATTCACCAACGGCAGATTCATCTGCTGAACCTTGACAACTCCTCGAGTGAC	180
QY	1068	AATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTTAGTTCCACC	1127
DB	181	AATAGTAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTTAGTTCCACC	240
QY	1128	TGCTATTGAAGTACGCTTCATCGATCGAAGGCTGTCCAGAAAGTTTATTACCTAAA	1187
DB	241	TGCTATTGAAGTACGCTTCATCGATCGAAGGCTGTCCAGAAAGTTTATTACCTAAA	300
QY	1188	TGATGATCTCATGTTTGGAGGATGCTGCCAGATGATTTTACAGTCACCTCCAAAGG	1247
DB	301	TGATGATCTCATGTTTGGAGGATGCTGCCAGATGATTTTACAGTCACCTCCAAAGG	360
QY	1248	CCAGAAGGTTTATTGACATGGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCCCTG	1307
DB	361	CCAGAAGGTTTATTGACATGGCTGTGCCAAACTGTGCCGAGGGCTGCCAGGTTCCCTG	420
QY	1308	GATTAAAGGATGCTATTGTGACAGGCTTGTAAATTTACGCTGCGATGGGATGGTGG	1367
DB	421	GATTAAAGGATGCTATTGTGACAGGCTTGTAAATTTACGCTGCGATGGGATGGTGG	480
QY	1368	GGATTGCTCTGGAACAGTGAGGAGTGCCTATATTGAGGAGGTGGAGGTACTGGGAG	1427
DB	481	GGATTGCTCTGGAACAGTGAGGAGTGCCTATATTGAGGAGGTGGAGGTACTGGGAG	540
QY	1428	TATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTCTACTGTAA	1487
DB	541	TATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTTCTACTGTAA	600
QY	1488	TCAGGGATGCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGCTTT	1547
DB	601	TCAGGGATGCGGAATTCCTGGCTCGCTGATAGTTCTGTGACCAAGCATGCAATGCTTT	660
QY	1548	GTCCTGTGGGTTTGATGCTGCGACTGTGGCAAGATCATTTTCATGAATGTATAAGT	1607
DB	661	GTCCTGTGGGTTTGATGCTGCGACTGTGGCAAGATCATTTTCATGAATGTATAAGT	720
QY	1608	GATCCTTCTCCCAACACAGACTCACTATATTTCAAAAAGGTGAATGCTGCTTATTTT	1667
DB	721	GATCCTTCTCCCAACACAGACTCACTATATTTCAAAAAGGTGAATGCTGCTTATTTT	780
QY	1668	CAGCTTTGCAAGATAGCCAAAGAGAGTGAAGGTGCTGTATAGTACCAATCCAATAAT	1727
DB	781	CAGCTTTGCAAGATAGCCAAAGAGAGTGAAGGTGCTGTATAGTACCAATCCAATAAT	840
QY	1728	TCGACATGCTTCTATTGCCAACAAAGTGGAAACCATCTCACCTCATATATGACAGTGGAA	1787
DB		TCGACATGCTTCTATTGCCAACAAAGTGGAAACCATCTCACCTCATATATGACAGTGGAA	

DB	841	TCGACATGCTTCTATTGGCCAAAGTGGAAACCATCTCACCTCATATAATGCACAGTGGAA	900
QY	1788	GAATGCCACCAATACATTTAAATCTCAGCTTCAAAATACAAAGCATGAAGAGTCAA	1847
DB	901	GAATGCCACCAATACATTTAAATCTCAGCTTCAAAATACAAAGCATGAAGAGTCAA	960
QY	1848	AATGSCAGATAACAGTGGAGGTGGACACAAAGGAGGACCAAAACCTGAATCTTACGCGCCA	1907
DB	961	AATGSCAGATAACAGTGGAGGTGGACACAAAGGAGGACCAAAACCTGAATCTTACAGCCA	1020
QY	1908	GAAGGTTACGAAATTTAGTTAGTCCCAATAACATCTTCCAGAGGCGGAAATCCTTTT	1967
DB	1021	GAAGGTTACGAAATTTAGTTAGTCCCAATAACATCTTCCAGAGGCGGAAATCCTTTT	1080
QY	1968	TGAGGATATTCCCAAGAAAAAGCGTTCCCGAAGTTTAAAGACACATGATGTTAACTCAAC	2027
DB	1081	TGAGGATATTCCCAAGAAAAAGCGTTCCCGAAGTTTAAAGACACATGATGTTAACTCAAC	1140
QY	2028	AAGGAGAGCCAGGAGAGGTGAAATTTCCCTGGTAAATATTTCACCTCTTCCAAAAA	2087
DB	1141	AAGGAGAGCCAGGAGAGGTGAAATTTCCCTGGTAAATATTTCACCTCTTCCAAAAA	1200
QY	2088	CGCCCAAGTTGAGTCTCAATACCTTGGATTGCAACTGGAACATGGAGACATCCTTTGAA	2147
DB	1201	CGCCCAAGTTGAGTCTCAATACCTTGGATTGCAACTGGAACATGGAGACATCCTTTGAA	1260
QY	2148	AGGATACAAATTTGTCAGCTCAGCCTTGTGATGATCATTTCTGTGAACTCACACATGC	2207
DB	1261	AGGATACAAATTTGTCAGCTCAGCCTTGTGATGATCATTTCTGTGAACTCACACATGC	1320
QY	2208	TAAATTAATAATCAAGCTTAAATTAACAGATGAAACAAATGACAGTCTTGGTGGCTCCACA	2267
DB	1321	TAAATTAATAATCAAGCTTAAATTAACAGATGAAACAAATGACAGTCTTGGTGGCTCCACA	1380
QY	2268	GGAAAAACAGGTTCAATAAAGCATCTTGCACAAACAGCTTAGGAGTGTCTGAAAGATGCA	2327
DB	1381	GGAAAAACAGGTTCAATAAAGCATCTTGCACAAACAGCTTAGGAGTGTCTGAAAGATGCA	1440
QY	2328	GAGGTTGACTTTTCCCTGCGAGTGTAAAAGTGAATGGTATGACACAGGCTCAGAAATCC	2387
DB	1441	GAGGTTGACTTTTCCCTGCGAGTGTAAAAGTGAATGGTATGACACAGGCTCAGAAATCC	1500
QY	2388	ACCCCTGGACTTTGGAGACACAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCAT	2447
DB	1501	ACCCCTGGACTTTGGAGACACAGCAAGATTTAGAGTGGAACTCACACCCAAAAACCAT	1560
QY	2448	AGGCGGAAATGTGACAAAAGAAAAAGCCCATCTCTGATTTGCCACTTGGAAAGCCAGAT	2507
DB	1561	AGGCGGAAATGTGACAAAAGAAAAAGCCCATCTCTGATTTGCCACTTGGAAAGCCAGAT	1620
QY	2508	GACAAAAGAAAAAATACAGGGAAGAAAAAGAGAACAGTAGAATGGAGGAAAAATGC	2567
DB	1621	GACAAAAGAAAAAATACAGGGAAGAAAAAGAGAACAGTAGAATGGAGGAAAAATGC	1680
QY	2568	TGAAAATACATAGGCGTTTACTGAAAGTGTACTTGGAGAGAAAGCTGACAGATTTACACAGA	2627
DB	1681	TGAAAATACATAGGCGTTTACTGAAAGTGTACTTGGAGAGAAAGCTGACAGATTTACACAGA	1740
QY	2628	TAGTTACTTGGGCTTTTGGCAATGGAGAAAAAAGTATTTCCCTAGATCTTCCGAGCA	2687
DB	1741	TAGTTACTTGGGCTTTTGGCAATGGAGAAAAAAGTATTTCCCTAGATCTTCCGAGCA	1800
QY	2688	AGAAGAGCATTTGAAGACACAAATTTGGCCTACTTCACTGATAGCAAGAAATAGAGCCAGATA	2747
DB	1801	AGAAGAGCATTTGAAGACACAAATTTGGCCTACTTCACTGATAGCAAGAAATAGAGCCAGATA	1860
QY	2748	CAAGAGAGATACATTTGAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGTT	2807
DB	1861	CAAGAGAGATACATTTGAGATTCCTCAGATATGTAATAAATTTCTAAATAGCAAGTT	1920
QY	2808	TGATTTACATCGCGGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGAGTTGTTAT	2867
DB	1921	TGATTTACATCGCGGAAAGTCCCTGCTCACATGCTCACATGATTTGACCGAGTTGTTAT	1980



QY 2868 GCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGACAAGACGCTCATTTACAAAGTGG 2927  
 Db 1981 GCAAGAACTGCAAGATATGTTCCCTGAGAAATTTGACAAGACGCTCATTTACAAAGTGG 2040  
 QY 2928 CCATTCGAGGATGAGTGGTTCCTCTCTCTATTTTATTTATCTCATGAGTGCAGTGA 2987  
 Db 2041 CCATTCGAGGATGAGTGGTTCCTCTCTCTATTTTATTTATCTCATGAGTGCAGTGA 2100  
 QY 2988 GCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTTGTC 3047  
 Db 2101 GCCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTCTTTGTC 2160  
 QY 3048 TGACAGAGAATCCGAACACTGCTACCAAGTTCAGAACTCCGCTTAAGTTTGAGGA 3107  
 Db 2161 TGACAGAGAATCCGAACACTGCTACCAAGTTCAGAACTCCGCTTAAGTTTGAGGA 2220  
 QY 3108 TTTGACAGGCTGGAACACATGCTAATAAATGCTCAAAAATGCTTCTCTGATATCAC 3167  
 Db 2221 TTTGACAGGCTGGAACACATGCTAATAAATGCTCAAAAATGCTTCTCTGATATCAC 2280  
 QY 3168 GCAGCTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACGGT 3227  
 Db 2281 GCAGCTAAATAATTTCCACCACTCAGGAATCCTACTATGATCCCAACCTGCCACGGT 2340  
 QY 3228 CACTAAAGTCTAGTAACAACCTTAACACAGTAAGTACAAATCCACAAACATATTA 3287  
 Db 2341 CACTAAAGTCTAGTAACAACCTTAACACAGTAAGTACAAATCCACAAACATATTA 2400  
 QY 3288 GGACAAAACAAATATAGTTTGAATCATGCGGAGGAAGAAATCGCTTTTAAATGAT 3347  
 Db 2401 GGACAAAACAAATATAGTTTGAATCATGCGGAGGAAGAAATCGCTTTTAAATGAT 2460  
 QY 3348 TCGTACCAAGCTTTCTCATGTTGGTGGCCAGTTGGATGACATGAAGAAAACCTTAGGA 3407  
 Db 2461 TCGTACCAAGCTTTCTCATGTTGGTGGCCAGTTGGATGACATGAAGAAAACCTTAGGA 2520  
 QY 3408 GTTTGTTTGGCTGATGACACATGACCAATCATAAGATGCTCAGACAGTGAAGC 3467  
 Db 2521 GTTTGTTTGGCTGATGACACATGACCAATCATAAGATGCTCAGACAGTGAAGC 2580  
 QY 3468 TGTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAAG 3527  
 Db 2581 TGTCTCAGGACCTTCTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAAG 2640  
 QY 3528 AGAGTATCGAAACCGTTTCTTCATATGATGAGTGCAGGAATGAGGCGTTATCGGA 3587  
 Db 2641 AGAGTATCGAAACCGTTTCTTCATATGATGAGTGCAGGAATGAGGCGTTATCGGA 2700  
 QY 3588 CAAATTTGAAGT 3598  
 Db 2701 CAAATTTGAAGT 2711

RESULT 4  
 AK056137  
 LOCUS AK056137 2603 bp mRNA linear PRI 01-AUG-2002  
 DEFINITION Homo sapiens cDNA FLJ31575 fis, clone NT2R12001846, moderately similar to Basic domain/leucine zipper transcription factor.  
 VERSION AK056137  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens teratocarcinoma cell\_line:NT2 cDNA to mRNA, clone\_lib:NT2R12 clone:NT2R12001846.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,

Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 2603)  
 Isogai,T., Otsuki,T. and Sugiyama,T.  
 Direct Submission  
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

## FEATURES

## source

Location/Qualifiers  
 1..2603  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2R12001846"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2R12"  
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"  
 64..>2603  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB71102.1"  
 /db\_xref="GI:16551459"  
 /translation="MLPKLQRTVTCLSHRYGLYVCLGVVTVIVSAFOGFEVLEW SRDQHYVLFDSYRONIAGSKSEONBCLCLPMPIDVVYTWVNGTDLELLKELQVRQMEP EQKAREILGNRTTEPYKSKSEKLECLLTHCIVKPMVLVDLPALNTILKDLPSLYS FHSASDIPNVAKPKNSTNVVDFSTKDVEDAHLLGNSTQTVWRGTLTIDKEV PCLVLMQDLAEISGFPTFKETNQLKLPENLSSKVKLLQLYSEAVLALKNNPKD POELNKOTKNNITDGLKELTISPAYLWLDLDSAISQSKODEDISASREFDNBELYSLR SIERHAPWNRNIFIVTNGOIPSWNLNDNPRVTIVTHODVFNLSHLTFFSPGATESHI HRIEGLSOKETLYLNDUMFGKDVMPDDDFYSHSKGKYVLTVPVPCNCAEGCPGSKIDG YCDKRNNSACDWDGDCSGNSGSRITAGGGGSGVGPWOPFGGSGINSYSCNQG CANSLWADKFCDCACNVLSGCFDAGDCQDHFHELYKILLPNQTHLILPAGCELPITF SFAEYAKRGVEGAYSDNPITRHASTANKWTIHLHSGMNNATTIHPNLTPQNTNDEE FKMQITTEVDTRGPKLNSTAGKYENLVSPITLLPEAEILFIDIPKEKFPKFRHD VNSTRRAQEEVKIPLNLSLLPKDQLSNTLDLQLEHGDITLKGYNLSKALLRSFL MNSOHAKTKNQAIITDEFNDSLVAPOEKQVHKSLPNSLGVSRQLRITRPVAVSKVN GHGQGNPPLDLETTARPERVETHQTGTIGGNVTKETKPPSLIVPLESOMT"

BASE COUNT 779 a 553 c 616 g 655 t  
 ORIGIN

Query Match 66.8%; Score 2404.8; DB 9; Length 2603;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2406; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 103 AGCCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGAGACAATATTGCTGGAAGTCC 162  
 Db 196 AGCCGAGATCAATACCATGTTTGTGTTTGTATTCCTATAGAGACAATATTGCTGGAAGTCC 255  
 QY 163 TTTCAGAAATCGGCTTGTCTGCCCATGCCGATTGACGTTGTTTACACCTGGTGAATGCC 222  
 Db 256 TTTCAGAAATCGGCTTGTCTGCCCATGCCGATTGACGTTGTTTACACCTGGTGAATGCC 315  
 QY 223 ACAGATCTTGAATCTACTAGGAACACTACAGAGGTTCAGAGAACAGATGGAGGAGGACG 282  
 Db 316 ACAGATCTTGAATCTACTAGGAACACTACAGAGGTTCAGAGAACAGATGGAGGAGGACG 375  
 QY 283 AAAGCAATGAGAAATCTCTGGGAAAAACACACGGAACCTACTTAAGAAGAGTGAAGA 342  
 Db 376 AAAGCAATGAGAAATCTCTGGGAAAAACACACGGAACCTACTTAAGAAGAGTGAAGA 435

QY 343 CAGTTAGAGTGTCTTAAACACTGCAATTAAGGTGCCAATGCTTGCTGGACCCAGCC 402  
DB 436 CAGTTAGAGTGTCTTAAACACTGCAATTAAGGTGCCAATGCTTGCTGGACCCAGCC 495  
QY 403 CTGCCAGCCACATCACCCTCAAGGACCTGCCATCTCTTTATCCCTCTTTTCATCTGCCC 462  
DB 496 CTGCCAGCCACATCACCCTCAAGGACCTGCCATCTCTTTATCCCTCTTTTCATCTGCCC 555  
QY 463 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTT 522  
DB 556 AGTGACATTTCAATGTTGCAAAACCAAAACCCCTCTACCAATGCTCAGTTGTTGTT 615  
QY 523 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTCGACTCTCTTAAAGGAAATAGCAGA 582  
DB 616 TTTGACAGTACTAAGGATGTTGAAGATGCCACTCTCGACTCTCTTAAAGGAAATAGCAGA 675  
QY 583 CAGACAGTATGGAGGGCTACTTGCACACAGATAAAGAAAGTCCCTGGATTAGTCTAATG 642  
DB 676 CAGACAGTATGGAGGGCTACTTGCACACAGATAAAGAAAGTCCCTGGATTAGTCTAATG 735  
QY 643 CAAGATTTGGCTTCCCTGAGTGATTTCCACCAACATTCAGGAAACAAATCAACTAAAA 702  
DB 736 CAAGATTTGGCTTCCCTGAGTGATTTCCACCAACATTCAGGAAACAAATCAACTAAAA 795  
QY 703 ACAAATTTGCCAGAAATCTTCTCTTAAAGTCAAACTGTGCGAGTTGATTCAGAGGCC 762  
DB 796 ACAAATTTGCCAGAAATCTTCTCTTAAAGTCAAACTGTGCGAGTTGATTCAGAGGCC 855  
QY 763 AGTGTACGGCTTCTAAACTGATTAACCCCAAGGATTTCAAGAAATGAATGAACAACT 822  
DB 856 AGTGTACGGCTTCTAAACTGATTAACCCCAAGGATTTCAAGAAATGAATGAACAACT 915  
QY 823 AAGAAGAACATGACCATTTGATGAAAGAACTGACCATGAGTCTGCATTTATATGCG 882  
DB 916 AAGAAGAACATGACCATTTGATGAAAGAACTGACCATGAGTCTGCATTTATATGCG 975  
QY 883 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGGTTTGA 942  
DB 976 GATCTGAGGCCATCAGCCAGTCTAAGCAGGATGAAGACATCTCTGCCAGTGGTTTGA 1035  
QY 943 GATAACGAAGAACTGAGTACTCATTTGCCATCTATCGAGAGGATGACCATGGTTCGG 1002  
DB 1036 GATAACGAAGAACTGAGTACTCATTTGCCATCTATCGAGAGGATGACCATGGTTCGG 1095  
QY 1003 AATATTTTTCATTTGTCACCAACGGGAGATTCATCTGCTCAACTGACAACTCCTCGA 1062  
DB 1096 AATATTTTTCATTTGTCACCAACGGGAGATTCATCTGCTCAACTGACAACTCCTCGA 1155  
QY 1063 GTGACAAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTTAGT 1122  
DB 1156 GTGACAAATAGTAAACACACAGGATGTTTTCGAAATTTGAGCCACTTGCTACCTTTAGT 1215  
QY 1123 TCACCTGCTATTGAAAGTCACTGTCATCGATCGAAGGGCTGTCCAGAAAGTTTATTTAC 1182  
DB 1216 TCACCTGCTATTGAAAGTCACTGTCATCGATCGAAGGGCTGTCCAGAAAGTTTATTTAC 1275  
QY 1183 CTAAATGATGATGTCATGTTTGGGAAGGATGCTTGCCAGATGATTTTACAGTCACTCC 1242  
DB 1276 CTAAATGATGATGTCATGTTTGGGAAGGATGCTTGCCAGATGATTTTACAGTCACTCC 1335  
QY 1243 AAAGGCCAGAAAGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGT 1302  
DB 1336 AAAGGCCAGAAAGTTTATTTGACATGGCTGTGCCAAACTGTGCCAGGGCTGCCAGGT 1395  
QY 1303 TCCTGGATTAGGATGGCTATTGTGACAAGGCTTGTATAATTCAGGCTGCGATTGGGAT 1362  
DB 1396 TCCTGGATTAGGATGGCTATTGTGACAAGGCTTGTATAATTCAGGCTGCGATTGGGAT 1455  
QY 1363 GGTGGGGATTGCTCTGGAAACAGTGGAGGGATGCGCTATATTGACAGAGGTGGAGTACT 1422  
DB 1456 GGTGGGGATTGCTCTGGAAACAGTGGAGGGATGCGCTATATTGACAGAGGTGGAGTACT 1515

QY 1423 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1482  
DB 1516 GGGAGTATTGGAGTTGGACAGCCCTGGCAGTTTGGTGGAGGAATAAACAGTGTCTCTTAC 1575  
QY 1483 TGTAAATCAGGGATGCGGAATTCCTGGCTCGCTGATGAAGTTCTGTGACCAAGCATCAAT 1542  
DB 1576 TGTAAATCAGGGATGCGGAATTCCTGGCTCGCTGATGAAGTTCTGTGACCAAGCATCAAT 1635  
QY 1543 GTCTTCTGCTGTGGTGTGATGCTGCGACTGTGGCAGAGATCATTTTCATGAATTTGAT 1602  
DB 1636 GTCTTCTGCTGTGGTGTGATGCTGCGACTGTGGCAGAGATCATTTTCATGAATTTGAT 1695  
QY 1603 AAAGTATGCTTCTCCCAAAACAGACTCACTATATTTATTTCCAAAAGTGAATGCTGCT 1662  
DB 1696 AAAGTATGCTTCTCCCAAAACAGACTCACTATATTTATTTCCAAAAGTGAATGCTGCT 1755  
QY 1663 TATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGAAGTGGCTATAGTGACAATCCA 1722  
DB 1756 TATTTTCAGCTTTGCGAAGTAGCCAAAAGAGAGTTGAAGTGGCTATAGTGACAATCCA 1815  
QY 1723 ATAATTCGACATGCTTCTATTGGCCAAAGTGGAAAACCATCCACCTCATATGCACAGT 1782  
DB 1816 ATAATTCGACATGCTTCTATTGGCCAAAGTGGAAAACCATCCACCTCATATGCACAGT 1875  
QY 1783 GGAATGAATGCCACACAAATACATTTTAACTCTCAGTTTCAAAAATACAAACGATGAAG 1842  
DB 1876 GGAATGAATGCCACACAAATACATTTTAACTCTCAGTTTCAAAAATACAAACGATGAAG 1935  
QY 1843 TTCAAAATTCGACATGAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACG 1902  
DB 1936 TTCAAAATTCGACATGAGTGGAGTGGACACAAAGGAGGACCAAACTGAATTTCTACA 1995  
QY 1903 GCCCAGAAAGGTTACGAAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGCGGAAATC 1962  
DB 1996 GCCCAGAAAGGTTACGAAAATTTAGTTAGTCCCATTAACACTTCTTCAGAGGCGGAAATC 2055  
QY 1963 CTTTTTGAGGATATTCCTCAAGAAAACGGTTCCCGAAGTTTAAAGACATGATGTTAAC 2022  
DB 2056 CTTTTTGAGGATATTCCTCAAGAAAACGGTTCCCGAAGTTTAAAGACATGATGTTAAC 2115  
QY 2023 TCAACAAGGAGAGCCCGAGGAGAGTGAATAATCCCTCGTGAATATTTTACCTCTTCCA 2082  
DB 2116 TCAACAAGGAGAGCCCGAGGAGAGTGAATAATCCCTCGTGAATATTTTACCTCTTCCA 2175  
QY 2083 AAAGACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACT 2142  
DB 2176 AAAGACGCCAGTGTAGTCTCAATACCTTGGATTTGCAACTGGAACATGAGACATCACT 2235  
QY 2143 TTGAAAGGATACAAATTTGCTCAAGTCAAGCTTGCCTGAGATCATTTCTGATGAATCAGAG 2202  
DB 2236 TTGAAAGGATACAAATTTGCTCAAGTCAAGCTTGCCTGAGATCATTTCTGATGAATCAGAG 2295  
QY 2203 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATCAGAGTTTGGTGGCT 2262  
DB 2296 CATGCTAAAATAAAAATCAAGCTATAATAACAGATGAACAAATCAGAGTTTGGTGGCT 2355  
QY 2263 CCACAGGAAAACAGGTTTCATAAAGCATCTTGCCAAAACAGCTTAGAGTGTCTGAAAGA 2322  
DB 2356 CCACAGGAAAACAGGTTTCATAAAGCATCTTGCCAAAACAGCTTAGAGTGTCTGAAAGA 2415  
QY 2323 TTGACAGGTTGACTTTTCTGCAAGTGAAGTGAATGGTCAATGAGGCTCAG 2382  
DB 2416 TTGACAGGTTGACTTTTCTGCAAGTGAAGTGAATGGTCAATGAGGCTCAG 2475  
QY 2383 AATCCACCCCTGGACTGGAGACACAGCAAGATTTAGAGTGAACACTCACCACCAAAA 2442  
DB 2476 AATCCACCCCTGGACTGGAGACACAGCAAGATTTAGAGTGAACACTCACCACCAAAA 2535  
QY 2443 ACCATAGGGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTCTTCCACTGGAAAGC 2502  
DB 2536 ACCATAGGGGAAATGTGACAAAAGAAAAGCCCATCTCTGATTTCTTCCACTGGAAAGC 2595  
QY 2503 CAGATGAC 2510

Db 2596 CAGATGAC 2603  
|||||||

## RESULT 5

AC010205

LOCUS

DEFINITION

AC010205

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC010205 162427 bp DNA linear PRI 13-AUG-2002  
Homo sapiens 12 BAC RP11-285E23 (Roswell Park Cancer Institute  
Human BAC Library) complete sequence.

AC010205  
AC010205.5 GI:6468049  
HTG.  
human.

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 162427)

Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,

Bodota, B., Bock, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,

Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,

David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,

Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,

Forcum-Ransey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,

Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M.,

Holloway, C., Hosak, H., Issar, A., Jackson, L.E., Jackson, L., Jia, Y.,

Jones, M., Kelly, S., Kneitz, S., Kondejowski, N., Kong, Y., Kovar, C.,

Lau, S., Leal, B., Lee, E., Li, Z., Lichtarge, O., Liu, J., Liu, W.,

Logan, O., Lu, J., Lucier, R., Marondel, I., Martin, R., Martinez, C.,

McLeod, M.P., Mei, G., Merscher, S., Miller, A., Montgomery, K.T.,

Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N.,

Nguyen, S., Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L.,

Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J.,

Scherer, S., Shah, E., Shen, H., Shim, C., Simon, M., Sparks, A.,

Stamps, A., Sugang, R., Tabor, P., Taylor, R., Vasquez, L., Vinson, R.,

Vo, Q., Wabwah, M., Watlington, S., Weinstein, G., Weinstein, I.R.,

Williamson, A., Worley, K., Wren, J., Wrenford, G., Xiang, A.M.,

Yang, R., Yu, W., Zhou, X., Kuchelapatti, R., Nelson, D. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 162427)

Worley, K.C.

Direct Submission

Submitted (15-SEP-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 162427)

Worley, K.C.

Direct Submission

Submitted (25-NOV-1999) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 162427)

Worley, K.C.

Direct Submission

Submitted (13-AUG-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 25, 1999 this sequence version replaced gi:6087851.

INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email

gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the

entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the

Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches

of a local database that includes entries from dbSTS, GDB, and

local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAT-REPORT-----

## ----- Summary Statistics -----

Contig length: 144841

Phrap values in estimate: 143745

Average error rate (BCM-Phrap estimate): 0.000323023

Fraction of Phrap values less than 40 : 0.052607

Number of consensus changing edits: 22

Number of N's in consensus : 0

## ----- Consensus changing edits -----

Position	Original+Context	Edited+Context
1530	atataatgaaa(n)tgagtagac	atataatgaaa(t)tgagtagac
3935	ctaaattgta(n)tgatatttt	ctaaattgta(t)tgatatttt
5613	tggtttttg(n)ttctgttgtt	tggtttttg(t)ttctgttgtt
25390	ggctttattg(n)aattatgaat	ggctttattg(g)aattatgaat
47091	ggtagaaaatt(n)ccctgtgtaa	ggtagaaaatt(c)ccctgtgtaa
65289	taatgcagca(t)tatacctct	taatgcagca(a)tatacctct
98027	gtcacagaag(n)ccacaatgta	gtcacagaag(t)ccacaatgta
105124	tctaatttaa(n)caagggaagt	tctaatttaa(c)caagggaagt
105232	aacagaatac(n)agacactggg	aacagaatac(t)agacactggg
105335	gtgcagct(n)agatnncactn	gtgcagct(g)agatnncactn
105340	atgctnagat(n)catnnnctc	atgctnagat(c)catnnnctc
105345	agatnncact(n)nctcactgg	agatnncact(g)actcactgg
105346	agatnncact(n)nctcactgg	agatnncact(a)actcactgg
105347	gatacactnn(n)ctcactggac	gatacactga(a)ctcactggac
105371	ttcacttac(n)nnnnaaaat	ttcacttac(a)gctcnaaaat
105372	ttcacttac(n)nnnnaaaat	ttcacttac(g)gctcnaaaat
105373	ttcacttac(n)nnnnaaaat	ttcacttac(c)gctcnaaaat
105374	ttcacttac(n)nnnnaaaat	ttcacttac(c)gctcnaaaat
105376	ttcacttac(n)nnnnaaaat	ttcacttac(c)gctcnaaaat
105503	ccctattctt(n)atattagaat	ccctattctt(c)atattagaat
105569	ttttttttt(n)cttgagacag	ttttttttt(t)cttgagacag
112885	tttattaaga(n)atataataa	tttattaaga(a)atataataa

## ----- Distribution of Quality &lt; 40 Bases -----

# bases	* * * * *
10001	* * * * *
9001	* * * * *
8001	* * * * *
7001	* * * * *
6001	* * * * *
5001	* * * * *
4001	* * * * *
3001	* * * * *
2001	* * * * *
1001	* * * * *
01	* * * * *



BAC Library) complete sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

AC063950  
AC063950.37 GI:18390143

HTG.  
Homo sapiens  
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS

1 (bases 1 to 174231)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayelle, M., Banks, P.,  
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brivava, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S.,  
Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,  
Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,  
Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P.,  
Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M.,  
Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S.,  
Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y.,  
Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U.,  
King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N.,  
Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O.,  
Lieu, C., Liu, J., Liu, W., Loulsegue, H., Lozano, R.J., Lu, X.,  
Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,  
Marandel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,  
Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S.,  
Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,  
Mottabatt, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M.,  
Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, J.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojuboan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shim, C., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kucherlapati, R.,  
Weinstock, G. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Unpublished  
2 (bases 1 to 174231)  
Worley, K.C.  
Direct Submission  
Submitted (22-APR-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

3 (bases 1 to 174231)  
Worley, K.C.  
Direct Submission  
Submitted (29-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jan 29, 2002 this sequence version replaced gi:17149338.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.

Overlapping clones are noted at the beginning and end of the  
Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES	Source	Location/Qualifiers
STS	1..174231	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="12"
		/clones="RP11-511H9"
	132..402	/standard_name="133116"
repeat_region	626..927	/rpt_family="AluSx"
repeat_region	1405..1699	/rpt_family="AluSg"
repeat_region	2197..2330	/rpt_family="AluJb"
repeat_region	2334..2375	/rpt_family="A-rich"
repeat_region	3303..3346	/rpt_family="AT-rich"
repeat_region	3561..3860	/rpt_family="AluSx"
repeat_region	4930..4957	/rpt_family="AT-rich"
STS	6296..6422	/standard_name="97243"
STS	6356..6479	/standard_name="48565"
repeat_region	6607..6638	/rpt_family="AT-rich"
repeat_region	7063..7087	/rpt_family="AT-rich"
repeat_region	7594..7889	/rpt_family="AluSx"
STS	8562..8691	/standard_name="46931"
STS	8777..8893	/standard_name="3096"
repeat_region	complement(9277..9424)	/rpt_family="L2"
repeat_region	complement(9693..9967)	/rpt_family="L2"
repeat_region	complement(10118..10410)	/rpt_family="AluJb"

```
repeat_region 11336..11363
/rpt_family="AT_rich"
repeat_region 11518..11700
/rpt_family="L2"
repeat_region complement(11714..11865)
/rpt_family="MIR"
repeat_region 11876..12009
/rpt_family="L2"
repeat_region 12057..12280
/rpt_family="MER2"
repeat_region complement(12281..12596)
/rpt_family="AluSx"
repeat_region 12597..12668
/rpt_family="MER2"
repeat_region 12671..12973
/rpt_family="MLT1K"
repeat_region 13905..14038
/rpt_family="FLAM_C"
repeat_region complement(14220..14468)
/rpt_family="AluSx"
repeat_region 15007..15300
/rpt_family="AluSc"
repeat_region 15478..15498
/rpt_family="(TA)n"
repeat_region complement(15499..15630)
/rpt_family="FLAM_C"
repeat_region 16321..16367
/rpt_family="(TG)n"
repeat_region 16935..16955
/rpt_family="AT_rich"
repeat_region 17746..17767
/rpt_family="AT_rich"
repeat_region 17768..18079
/rpt_family="AluSq"
repeat_region 18081..18363
/rpt_family="AluSx"
repeat_region 19327..19625
/rpt_family="AluSx"
repeat_region 20207..20227
/rpt_family="AT_rich"
repeat_region complement(20293..20617)
/rpt_family="Aluub"
repeat_region 20901..20984
/rpt_family="MIR"
repeat_region 21019..21125
/rpt_family="L2"
repeat_region 21581..21698
/rpt_family="FLAM_C"
repeat_region complement(21754..21874)
/rpt_family="L2"
repeat_region complement(21902..21977)
/rpt_family="MIR"

Query Match 30.78; Score 1103.6; DB 9; Length 174231;
Best Local Similarity 99.6%; Pred. No. 4.8e-278;
Matches 1106; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1579 CAAGATCATTTTCATGAATTCATATAAGTGTATCTCTCCCAACACAGACTCCTATATT 1638
DB 44934 CTAGATCATTTTCATGAATTCATATAAGTGTATCTCTCCCAACACAGACTCCTATATT 44875

QY 1639 ATTCAAAAGGTGAATGCTCGCTTATTTCAGCTTTTCAGAAGTAGCCAAAAGAGAGTT 1698
DB 44874 ATTCAAAAGGTGAATGCTCGCTTATTTCAGCTTTTCAGAAGTAGCCAAAAGAGAGTT 44815

QY 1699 GAAGTGCCTATAGTACATCCCAATTAATTCGACATGCTTCTATTGGCACAAGTGAAA 1758
DB 44814 GAAGTGCCTATAGTACATCCCAATTAATTCGACATGCTTCTATTATTCGACAAGTGAAA 44755

QY 1759 ACCATCCACCTCATAAATGCACAGTGGGAATGAATGCCACCAATACATTTTAATCTCACG 1818
DB 44754 ACCATCCACCTCATAAATGCACAGTGGGAATGAATGCCACCAATACATTTTAATCTCACG 44695
```

```
QY 1819 TTTCAAATACAAACGATGAAGAGTTCAAATTCAGATAACAGTGGAGTGGACACAAGG 1878
DB 44694 TTTCAAATACAAACGATGAAGAGTTCAAATTCAGATAACAGTGGAGTGGACACAAGG 44635

QY 1879 GAGGACCAAACTGAATTCCTAGGCCCAAGAGGTTACGAAAATTTAGTTAGTCCATA 1938
DB 44634 GAGGACCAAACTGAATTCCTAGGCCCAAGAGGTTACGAAAATTTAGTTAGTCCATA 44575

QY 1939 ACACCTTCTCCAGAGCGGAAATCCCTTTTGGAGGATATTCCTCCAAAGAAAAGCTTCCG 1998
DB 44574 ACACCTTCTCCAGAGCGGAAATCCCTTTTGGAGGATATTCCTCCAAAGAAAAGCTTCCG 44515

QY 1999 AAGTTTAAAGAGCATGATGTTAACTCAACAAGAGAGAGCCAGAGAGGTGAAAATTCOC 2058
DB 44514 AAGTTTAAAGAGCATGATGTTAACTCAACAAGAGAGAGCCAGAGAGGTGAAAATTCOC 44455

QY 2059 CTGGTAAATATTTCACTCTCTTCCAAAAGACGCCAGTTCAGTCTCAATACCTTGGATTG 2118
DB 44454 CTGGTAAATATTTCACTCTCTTCCAAAAGACGCCAGTTCAGTCTCAATACCTTGGATTG 44395

QY 2119 CAATGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCCAAGTCAGCCTTGCTG 2178
DB 44394 CAATGGAACATGGAGACATCACTTTGAAAGGATACAATTTGTCCAAGTCAGCCTTGCTG 44335

QY 2179 AGATCATTTCTGATGAACATCAGACATGCTAAATAAAAAATCAAGCTATAATAACAGAT 2238
DB 44334 AGATCATTTCTGATGAACATCAGACATGCTAAATAAAAAATCAAGCTATAATAACAGAT 44275

QY 2239 GAAACAAATGACAGTTTGGTGGCTCCACAGGAAAACAGAGTTTCATAAAGCATCTTTGCCA 2298
DB 44274 GAAACAAATGACAGTTTGGTGGCTCCACAGGAAAACAGAGTTTCATAAAGCATCTTTGCCA 44215

QY 2299 AACAGCTTAGGAGTGCTGAAAGATTGCAGAGGTTGACATTTTCTCCAGTGAAGTATAAA 2358
DB 44214 AACAGCTTAGGAGTGCTGAAAGATTGCAGAGGTTGACATTTTCTCCAGTGAAGTATAAA 44155

QY 2359 GTGAATGGTCATGACACAGGCTCAGAAATCCACCCCTGGACTTGGAGACACACAGCAAGTTT 2418
DB 44154 GTGAATGGTCATGACACAGGCTCAGAAATCCACCCCTGGACTTGGAGACACACAGCAAGTTT 44095

QY 2419 AGATGGAACACTCACACCCAAAACCATAGGCGGAATGTGACAAAAGAAAGCCCGCA 2478
DB 44094 AGATGGAACACTCACACCCAAAACCATAGGCGGAATGTGACAAAAGAAAGCCCGCA 44035

QY 2479 TCTCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGAA 2538
DB 44034 TCTCTGATTGTTCCACTGGAAGCCAGATGACAAAAGAAAAGAAAATCACAGGGAAGAA 43975

QY 2539 AAAGAGACAGTAGAATGGAGGAAAATGCTGAAAATTCATATAGGCGTTTACTGAAGTGTA 2598
DB 43974 AAAGAGACAGTAGAATGGAGGAAAATGCTGAAAATTCATATAGGCGTTTACTGAAGTGTA 43915

QY 2599 CTGGGAAGAAAGCTGCAGATTACACAGATAGTTACTTGGGCTTTTGGCCATGGAGAAA 2658
DB 43914 CTGGGAAGAAAGCTGCAGATTACACAGATAGTTACTTGGGCTTTTGGCCATGGAGAAA 43855

QY 2659 AAAAAGTATTTCTAGATCTTCTCGACGAA 2688
DB 43854 AAAAAGTATTTCTAGATCTTCTCGACGTA 43825

RESULT 7
AC005409/c 177364 bp DNA linear PRI 27-MAY-2000
LOCUS Homo sapiens chromosome 10 clone 14951, complete sequence.
DEFINITION AC005409
VERSION AC005409.1 GI:4249432
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 177364)
```

AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Homo sapiens Chromosome 10 BAC Clone 14951 In I-CELL Disease Region  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-1998) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1999) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 4 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-1999) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 5 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-1999) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 6 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (27-FEB-1999) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 7 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAR-1999) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 8 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAY-2000) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 9 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAY-2000) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 10 (bases 1 to 177364)  
AUTHORS Fu, Y., Pan, H., D'Souza, A., Canfield, B. and Roe, B.A.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAY-2000) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Feb 9, 1999 this sequence version replaced gi:4235169.  
FEATURES  
source  
1. .177364  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="14951"  
BASE COUNT 53772 a 37167 c 35988 g 50437 t  
ORIGIN  
Query Match 30.6%; Score 1103.2; DB 9; Length 177364;  
Best Local Similarity 99.7%; Pred. NO. 6.le-278;  
Matches 1105; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1581 AGATCATTTTCATGAATGTATAAGTGATCCTTCTCCCAACACGAGCTCACTATATAT 1640  
|||||



```

RESULT 8
MUSBDLZ
LOCUS      1846 bp      mRNA      linear      ROD 02-MAY-1995
DEFINITION Mus Musculus basic domain/leucine zipper transcription factor mRNA,
3' end of cds.
ACCESSION L36434.1 GI:625041
VERSION    basic domain/leucine zipper transcription factor.
KEYWORDS   Mus musculus cDNA to mRNA.
SOURCE     Mus musculus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1846)
AUTHORS    Cordes, S.P. and Barsh, G.S.
TITLE      The mouse segmentation gene kr encodes a novel basic domain-leucine
zipper transcription factor
JOURNAL    Cell 79 (6), 1025-1034 (1994)
MEDLINE    95094266
PUBMED
FEATURES   Location/Qualifiers
            1..1846
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            1..1846
            <1..1155
            /codon_start=1
            /product="basic domain/leucine zipper transcription
            factor"
            /protein_id="AA65688.1"
            /db_xref="GI:625042"
            TGRFEVVKIPRVNLSLPEAOVRLSNLDLQERDITLKGYNLSKALLRFLGNS
            LDTKIKQARTDETKGLEVPQENPSRRPHGHRSEWTPAPETVTVKGRDHAL
            NPPVLNENRLEAGPTGLVTVSKENLSPLVPESHLPKEESDRAGNVPVKEIYP
            GRCSRILQAFCPKKKKYFQDLDAEESLKTOLAYFTDSKTKRGLKDFADSLRYV
            NKIUNSPGTSRVPAPMHMDIRIVMQLQDMFPEFDKTSFKRHVSEDMQFAFS
            YFYILMSAVOPLNISQVFEHVDYDQSVLSDREIRTLRIHLPLTCRT"
            1156..1846
            3'UTR
            polyA_site
            BASE COUNT 572 a 451 c 415 g 401 t 7 others
            ORIGIN
            Query Match 23.6%; Score 849.8; DB 10; Length 1846;
            Best Local Similarity 75.1%; Pred. No. 1.6e-211;
            Matches 1257; Conservative 0; Mismatches 341; Indels 75; Gaps 13;
            QY 1900 ACGGCCAGAGGGGTACGAAATTTAGTTAGTCCCATACACTCTCTCCAGAGCGCGGAA 1959
            Db 7 ACACCCAGAGGGCCTATGAAGTTTGGTTAGCCCGAGTGACACCTCTTCTCCAGCGCTGAC 66
            QY 1960 ATFCCTTTTGGAGGATATCCCAAGAAAACGCTTCCCGAAGTTTAAAGAGACATGATGTT 2019
            Db 67 GTCCCTTTTGAAGATGTCCTCCCAAGAGAAACGCTTCCCAAGATCAGAGACATGATGTA 126
            QY 2020 AACTCAACAGAGAGCGCCAGAGAGAGTGAATAATTCCTCGGTGAATAATTTCACTCCTT 2079
            Db 127 AATGCACAGGAGGATTCACAGAGAGGTGAAATTCCTCCCGGTGAATAATTTCACTCCTT 186
            QY 2080 CCAAAAGACGCGCTTTCAGTCTCAATACCTTGGATTTGCAACTGGCACTGGAGACATC 2139
            Db 187 CCCAAGAGGCCCGAGGTGAGGCTGAGCAACTTGGATTGGCACTAGAACGTGGAGACATC 246
            QY 2140 ACTTTGAAAGGATCAATTTTCCCAAGTCAGCGCTTGTGAGATCATTTCTTGATGAACATCA 2199
            Db 247 ACTCTGAAAGGATATACTTTGCCAAGTCAGCGCTTGTAGGTTCTTCTTGGGGAATTC A 306
            QY 2200 CAGCATGCTAAATAAATAACAGCTATATAACAGATGAACAAATACAGAGTTGGTG 2259
            Db 307 CTAGATATAAATAAATAACCTCAAGCT---AGGACCGATGAACAAAGGCAACCTGGAG 363
            QY 2260 GCTCCACAGGAAAAACAGGTTTCATAAAGATCTTGGCCAAACAGCTTAGGAGTCTCTGAA 2319

```

```

Db 364 GTCCACAGAGAAACCTTCTCCAGAGC-----TCCACATGGCTTGTGTGACAC 417
QY 2320 AGATTGCAGAGGTGACTTCTCTGCGAGTGTAAAAGTGAATGTCTATGACAGGGT 2379
Db 418 AGATCAGAGAGATGGACTGCCCCAGCAGACAGTGAACCGTGAAGCCGCTGACCAAGCT 477
QY 2380 CAGAAATCCACCCCTGGAGTTGGAGACACAGCAAGATTTAGAGTGGAAACTCACACCCA 2439
Db 478 TTGAATCCACCCCGGTGGTGGAGACCAATGCAAGATTGG-----CCGAG 522
QY 2440 AAAACCATAGCGGAATGTGACAAAAGAAAGCCGCCCTCTCTGATTTGTTCCACTGGAA 2499
Db 523 CCTACACTAGCGGTGACTGTGTCCAAAGAAACCTTTCACCGCTGATGCTGTCCCCAGAA 582
QY 2500 AGCCAGATGACAAAGAAAAGAAAATCACAGGGAAGAAAGAAAGAACAGTAGAATGGAG 2559
Db 583 AGCCACTTGCAAAAGAGAGAGAGTGCAGGGCAGAA----- 621
QY 2560 GAAATGCTGAAAATCACATAGGCGTTACTGAAGTGTACTTGGAAAGAAAGCTGCAGCAT 2619
Db 622 -----GGCAATGCTGTACCTGTAAGGAGTTAGTGCCTGGCAGCG----- 663
QY 2620 TACACAGATAGTTACTTGGGCTTTTGGCCATGGGAGAAAGAAATTTCTCTAGATCTT 2679
Db 664 TGCAGCAGAATTTATCCAGGCTTTTGGCCCTGGGAAAGAAAGATTTTCCAGACCTT 723
QY 2680 CTGCAGCAAGACAGTCATTTGAAGACACAATTTGGCCCTACTCTACATGATGACGAAGATA 2739
Db 724 CTTGTCGGGAAGAGTCATTTGAAGACCCAGTTTGGCCGTACTTTTACAGACAGCAACATACC 783
QY 2740 GCCAGATACAAAGAGATACATTTGCAGATTTCCCTCAGATATGTAAATAAATTTCTAAAT 2799
Db 784 GGGAGGCACTAAAAGATACATTTGCAGACTCCCTCCGATACGTCAATAAATTTCTCAAC 843
QY 2800 AGCAAGTTTGGATTACATCGCGGAAAGTCCCTGCTCAGATGCTCAGATGATTTGACCGG 2859
Db 844 AGCAAGTTTGGATTACATCCAGGAAAGTCCCTGTCACACATGCGGCACATGATTTGACAG 903
QY 2860 ATTGTTATGCAAGACTGCAAGATATGTTCCCTGGAAGATTTGACAAGACGTCATTTCAC 2919
Db 904 ATCGTTATGCAAGACTCCCAAGATATGTTCCCTGGAAGATTTGACAAGACTTCATTTCAC 963
QY 2920 AAGTGCAGCTTCTGAGGATATGAGTTTGGCTTCTCTTATTTTATTTATCTCATGAT 2979
Db 964 AAGTGCCTCCTCCTCAGGACATGCAAGTTTGGCTTCTCTCTACTTTTATTTACCTCATGAT 1023
QY 2980 GCAGTGCAGGCACTGAATATATCTCAAGTCTTTGATGAAGTTGATACAGATCAATCTGTT 3039
Db 1024 GCAGTTTCAAGCCCTCAATATTTCCCAAGTCTTTTCATGAAGTAGACACAGACCAATCTGT 1083
QY 3040 GTCTTCTCTGACAGAGAAATCCGAACACTGGCTACCAGAAATTCAGAACTGCGGTTAAGT 3099
Db 1084 GTCTTCTCTGATAGGAAATCCGAACCTGGCCACAGAAATTCACAGACCTACTCTTTAA-C 1142
QY 3100 TTGACAGGATTTGACAGGCTTGGAAACACATGCTAATAAATTTGC-TCAAAAATGCTTCTGTC 3158
Db 1143 TTGACAGGATTTGACAGGCTTGGAAACACATGTTAATAAATTTGCNNCAAAAATGCTCNCGCCG 1202
QY 3159 TGATATCAGCGACCTTAATAATTTCCACCACTCAGGAATCCTACTATGATGCCAACCT 3218
Db 1203 TAATAT-----CAACTCAACACATCCACCGACTCAGGAAGCATACTAGGACCCCAAGCT 1258
QY 3219 GCCACCGGT-CACATAAAGTCTAGTAACAACTGTAAACACTGTAAACACTGTACAAAATCCACA 3277
Db 1259 GCCTCGGTCCTACTAGAGTCTGTACACCACTGT--ACCAGTAACTGACAGATGCCACA 1316
QY 3278 AAGCATATAGGACAAAAACAATAATAGGTTTGAATCATGGGAGAAAGAAATTCGCTT 3337
Db 1317 AAGCCCTATAA-GACAAGAAACAATAACAGGTTTGAACCTCATGGGAGAGGAAGAAAGCTT 1375
QY 3338 TTAATAATGATTGCTACCAACGTTTCTCATGTGGTGGCCAGTTGGATGACATAAGAAAAA 3397

```



Db 1376 CAAGATGATACAAACCAAGTGTCTTCTCATGTGCTGGCCAGTTGGATGACATCAGAAAAA 1435

QY 3398 ACCTAGGAAGTTGTTGCTGAATGACAAACATTGACCACATCATAAAGATG-CTCAG 3456

Db 1436 ACCCCAGGAAGTTGTTGCTGAATGACAAAAATTGACCACAAACCATAAAGATGCCCGG 1495

QY 3457 ACAGTGAAGCTGTCTTCAGGGACTTCTATGAATCCATGTTCGCCATACCTTCCCAATTT 3516

Db 1496 ACAGTGAAGCTGTCTTCAGGGACTTCTATGAATCCATGTTCGCCATACCTTCCCAATTT 1555

QY 3517 GAAGTCCAGAGAGATGATGAAACCGTTTCCTTCATATGATGATGAGTGCAGGA 3569

Db 1556 GNGCTGCCAGAGAGTATCGAAACCGCTTTC-TGACATGATGATGAGTGCAGGA 1607

RESULT 9

HSM802690

LOCUS

DEFINITION Homo sapiens mRNA; cDNA DKFp762B226 (from clone DKFp762B226).

ACCESSION AL359588

VERSION AL359588.1 GI:8655649

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2428)

AUTHORS Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.

TITLE Direct Submission

JOURNAL Submitted (15-JUN-2000) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY

COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GfR (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFp762B226) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cdna/>.

## FEATURES

source

1..2428

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="496.6 cr from top of Chr12 linkage group"

/clone="DKFp762B226"

/tissue\_type="melanoma (MeWo cell line)"

/clone\_lib="762 (synonym: hmel2). Vector pSport1; host DH10B; sites NotI + SalI"

/dev\_stage="adult"

1..747

/gene="DKFp762B226"

<1..747

/gene="DKFp762B226"

/note="similarity to KIAA1208"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAB94874.1"

/db\_xref="GI:8655650"

/translations="PVRPVFVDVTDQSGVLSDRITLRIHELPLSLQDLTGLEH MLNCSKMLPADITQLNIPPTQESYDNLPRVTSLVNCKPQVTDKHAYKKNK YRFEIMGEIEAFKMTNIVSHVQGLDDIRKNPRFVCLINDINHDHKAQTVRAVL RDTYSMEFIPSPQFELPREYRNLHMELEQWRAYRDLKFLWCHVLTALIMFTIFS FFAEQIALRKRPFRRIHKEASPNRIY"

2411

polya\_site 725 a 483 c 460 g 760 t

BASE COUNT

ORIGIN

Query Match 16.5%; Score 592.8; DB 9; Length 2428;

Best Local Similarity 99.7%; Pred. No. 4.1e-144;

Matches 594; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3003 TCAAGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTGTCTGACAGAAATCCG 3062

Db 9 TCCGTCTTTGATGAAGTTGATACAGATCAATCTGGTGTCTGTCTGACAGAAATCCG 68

QY 3063 AACACTGGCTACCGAATTCACGAATCCCGTTAAAGTTTGAGGATTTGACAGTCTGGA 3122

Db 69 AACACTGGCTACCGAATTCACGAATCCCGTTAAAGTTTGAGGATTTGACAGTCTGGA 128

QY 3123 ACACATGCTAATAATTTGCTCAAAATGCTTCCTGCTGCTATATCAGCAGCTAAATATAT 3182

Db 129 ACACATGCTAATAATTTGCTCAAAATGCTTCCTGCTGCTATATCAGCAGCTAAATATAT 188

QY 3183 TCCACCACTCAGGAATCTCTACTATGATCCCAACCTGCCCGGTCACTAAAAGTCTAGT 3242

Db 189 TCCACCACTCAGGAATCTCTACTATGATCCCAACCTGCCCGGTCACTAAAAGTCTAGT 248

QY 3243 AACAACTGTAACACAGTAACATGACAAATCCACAAGCATATAGGACAAAAACAATA 3302

Db 249 AACAACTGTAACACAGTAACATGACAAATCCACAAGCATATAGGACAAAAACAATA 308

QY 3303 TAGGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAGTTTC 3362

Db 309 TAGGTTTGAATCATGGGAGAGAGAAATCGCTTTTAAATGATTCGTACCAAGTTTC 368

QY 3363 TCATGTGTTGGCCAGTTGGATGACATAGAAAAAACCCCTAGGAAGTTTGTTCCTGAA 3422

Db 369 TCATGTGTTGGCCAGTTGGATGACATAGAAAAAACCCCTAGGAAGTTTGTTCCTGAA 428

QY 3423 TGACAACATTTGACCACAAATCATAAAGATGCTCAGACAGTGAAGGCTGTTCAGGACTT 3482

Db 429 TGACAACATTTGACCACAAATCATAAAGATGCTCAGACAGTGAAGGCTGTTCAGGACTT 488

QY 3483 CTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAAGAGTAGTTCGAAACCG 3542

Db 489 CTATGAATCCATGTTCCCATACCTTCCCAATTTGAACTGCCAAGAGTAGTTCGAAACCG 548

QY 3543 TTCTCTTCATATGATGAGTGCAGGAATGGAGGCTTATCGAGACAAATGAAGT 3598

Db 549 TTCTCTTCATATGATGAGTGCAGGAATGGAGGCTTATCGAGACAAATGAAGT 604

RESULT 10

AC121402/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-324F15, \*\*\* SEQUENCING IN PROGRESS

AC121402

AC121402.2 GI:21903099

VERSION

KEYWORDS HTGS\_PHASE1.

SOURCE Rattus norvegicus.

ORGANISM Rattus norvegicus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 141215)

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayelle, M., Banks, T., Bartharia, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,

Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lounsged, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Whiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, G., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okuonu, G., Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 141215)  
 Worley, K. C.  
 Direct Submission  
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 141215)  
 Worley, K. C.  
 Direct Submission  
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 18, 2002 this sequence version replaced gi:20976340.

-----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GXAO  
 Center clone name: CH230-324F15  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 108630 bases at least Q40  
 Consensus quality: 111787 bases at least Q30  
 Consensus quality: 114494 bases at least Q20  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 42 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1223: contig of 1223 bp in length  
 1224 1323: gap of unknown length  
 1324 2695: contig of 1372 bp in length  
 2696 2795: gap of unknown length  
 2796 3942: contig of 1147 bp in length  
 3943 4042: gap of unknown length  
 4043 5211: contig of 1169 bp in length  
 5212 5312: gap of unknown length  
 5312 6586: contig of 1275 bp in length

6587 6687: gap of unknown length  
 6688 7910: contig of 1224 bp in length  
 7911 8011: gap of unknown length  
 8012 9051: contig of 1041 bp in length  
 9052 9151: gap of unknown length  
 9152 10546: contig of 1395 bp in length  
 10547 11871: contig of 1225 bp in length  
 11872 13687: contig of 1716 bp in length  
 13688 13788: gap of unknown length  
 13789 15781: contig of 1994 bp in length  
 15782 17949: contig of 2068 bp in length  
 17950 18050: gap of unknown length  
 18051 19781: contig of 1732 bp in length  
 19782 19881: gap of unknown length  
 19882 21439: contig of 1558 bp in length  
 21440 21539: gap of unknown length  
 21540 22816: contig of 1276 bp in length  
 22817 24651: contig of 1736 bp in length  
 24652 24751: gap of unknown length  
 24752 27211: contig of 2460 bp in length  
 27212 27311: gap of unknown length  
 27312 29806: contig of 2495 bp in length  
 29807 29906: gap of unknown length  
 29907 32159: contig of 2263 bp in length  
 32160 32269: gap of unknown length  
 32270 34289: contig of 2020 bp in length  
 34290 34389: gap of unknown length  
 34390 37438: contig of 3049 bp in length  
 37439 37538: gap of unknown length  
 37539 40323: contig of 2785 bp in length  
 40324 40423: gap of unknown length  
 40424 43380: contig of 2957 bp in length  
 43381 43480: gap of unknown length  
 43481 46912: contig of 3432 bp in length  
 46913 47012: gap of unknown length  
 47013 50378: contig of 3366 bp in length  
 50379 50478: gap of unknown length  
 50479 52969: contig of 2491 bp in length  
 52970 53069: gap of unknown length  
 53070 56475: contig of 3406 bp in length  
 56476 56575: gap of unknown length  
 56576 59055: contig of 2480 bp in length  
 59056 59155: gap of unknown length  
 59156 63855: contig of 4700 bp in length  
 63856 63955: gap of unknown length  
 63956 68690: contig of 4735 bp in length  
 68691 68790: gap of unknown length  
 68791 72439: contig of 3649 bp in length  
 72440 72539: gap of unknown length  
 72540 74492: contig of 1953 bp in length  
 74493 74592: gap of unknown length  
 74593 79801: contig of 5209 bp in length  
 79802 79901: gap of unknown length  
 79902 83391: contig of 3490 bp in length  
 83392 83491: gap of unknown length  
 83492 86836: contig of 3345 bp in length  
 86837 86936: gap of unknown length  
 86937 91723: contig of 4787 bp in length  
 91724 91823: gap of unknown length  
 91824 98085: contig of 6262 bp in length  
 98086 98185: gap of unknown length  
 98186 105770: contig of 7585 bp in length  
 105771 105870: gap of unknown length  
 105871 111548: contig of 5678 bp in length  
 111549 111648: gap of unknown length  
 111649 120177: contig of 8529 bp in length  
 120178 120277: gap of unknown length  
 120278 129704: contig of 9427 bp in length  
 129705 129804: gap of unknown length

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

\* 129805 141215: contig of 11411 bp in length.

FEATURES  
Source 1..141215  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-324F15"

BASE COUNT 38123 a 29118 c 28492 g 38666 t 6816 others  
ORIGIN

Query Match 13.98; Score 499; DB 2; Length 141215;  
Best Local Similarity 69.6%; Pred. No. 1.8e-119;  
Matches 777; Conservative 0; Mismatches 255; Indels 85; Gaps 4;  
QY 1572 CTGTGGGCAAGATCTTTCATGAATGTATAAAGTGCCTCTCCCTCCCAACACAGACTCA 1631  
DB 62137 CTTTCCTCTAGATCATTTTCATGAATGTATAAAGTGCCTCTCCCTCCCAACACAGACTCA 62078  
QY 1632 CTATATTATCCAAAGGTGAATGCCCTGCTTATTTTCAGCTTTTCGAGAAAGTAGCCAAAG 1691  
DB 62077 CTATGTTGTCCCAAGGTGAATACCTGTCTTATTTTCAGCTTTTCGAAACATAGCCAAAG 62018  
QY 1692 AGGAGTTCAAGGTGCCATATGACATCCATTAATTCGATGCTTCTATTGCCAAACAA 1751  
DB 62017 AGGGGTTGAGGGACCTACAGTGACAAACCCCATCGTCGTCACGCGTCTATCGCAACAA 61958  
QY 1752 GTGGAACACCATCCACCTCATTAATGCACAGTGGAAATGAATGCCACCAACATACATTTAA 1811  
DB 61957 GTGGAACACCATACCTCATTAATGCACAGTGGAAATGAATGCCACCAACATACATTTAA 61898  
QY 1812 TCTCAGCTTTCAAATCAACAGATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGA 1871  
DB 61897 CTTCACTTTCAAACCTCAACAGATGAAGAGTTCAAAATGCAGATAACAGTGGAGTGA 61838  
QY 1872 CACAGGAGGACCAACATGAATTCACGCCCCCAAGGTTTACGAAATTTAGTTAG 1931  
DB 61837 CACAGGAGGAGTGCCTAAACGAACTCCCAACCCAGAACCTATGACAAATTTGTTAG 61778  
QY 1932 TCCATAACACTCTTCCAGAGCGGAAATCTTTTTCAGATATTCCTCCCAAGAAACG 1991  
DB 61777 CCCAGTACACTCTTCTCAGGCTGAGTCCCTTTTGAAGATGTTCCCAAGAAACG 61718  
QY 1992 CTTCCCGAAGTTTAAAGACATGATGTTAACTCAACAGGAGAGCCCAAGAGAGTGA 2051  
DB 61717 CTTCCCGAGGTCAGGAGACCTGATGTAACGCAACAGGAAAGTCCAGGAGAGTGA 61658  
QY 2052 AATTCCCTGTAAATTTTCACTCTTCCAAAGAGCCCGCTGAGTCTCAATACCTT 2111  
DB 61657 AATCCCGGTGAATATTTCCCTCTTCCCAAGAGCCCGAGGTGAGGTGAGCAACCT 61598  
QY 2112 GGATTTGCACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCCAAAGTCAGC 2171  
DB 61597 GGATTTGCACTGGAACATGGAGACATCACTTTGAAAGGATACAAATTTGTCCAAAGTCAGC 61538  
QY 2172 CTTGCTGAGATCATTTCTGATGAACTCACAGCATGCTTAAATTAATAAATCAAGCTAAT 2231  
DB 61537 CTTGCTGAGATCATTTCTGATGAACTCACAGCATGCTTAAATTAATAAATCAAGCTAAT 61482  
QY 2232 AACAGATGAACAAATGACAGTTTGGTGGCTCCACAGAGAAACAGCTTTCATTAAGCAT 2291  
DB 61481 GACAGATGAACAAATGACAGTTTGGTGGCTCCACAGAGAAACAGCTTTCATTAAGCAT 61425  
QY 2292 CTTGCCAACAAGTTAGAGTGTCTGAAAGATTCAGAGGTGACTTTTCTTCAGTGAG 2351  
DB 61424 --TCCAGGTGCTTTGCTGCCGACCGACAGACAGAGGTGGACGACCCCGGAGAC 61368  
QY 2352 TGTAAGTGAATGGTCATGACAGGTGAGAAATCCACCCCTGGACTTGGAGACACAGC 2411  
DB 61367 AGCAGTGTGAACGGTGGTACACCGGTTTGAATCCACCCCGG----- 61325  
QY 2412 AAGATTAGATGGAACTCACACCCCAAAACCATAGGCGAAATGTGACAAAGAAA 2471  
DB 61324 -----ACCAACCTGTGCCCAAGAGAA 61302

QY 2472 GCCCCCATCTCTGATTGTTCCACTGGAAGCCAGATGACAAAGAAAGAAATCACAGG 2531  
DB 61301 CCTCTCACCACCTGATCATTTCCCCAGAGCCGATGCAAGAGGAGG----- 61253  
QY 2532 GAAAGAAAAAGAACACAGTAGAATGGAGAAAAATGCTGAAAAATCACATAGGCGTTACTGA 2591  
DB 61252 -----GACAGAGACAGACAGAGGCGCAATGCTGATAGCCAGGTACCTGTAAACGA 61203  
QY 2592 AGTGTTACTTGGGAAGAAAGCTGCAGCATTTACACAGATAGTTACTTGGGGCTTTTGGCCATG 2651  
DB 61202 GCTGGTGCCTGGCAGGAGCTACAGCAGCAGCATACCCAGTACCAGGCGTTTGGCCCTG 61143  
QY 2652 GGAGAAAAAAGTATTTCTAGATCTTCTCGACGAA 2688  
DB 61142 GGAGAAAAAAGTATTTCCAAAGACCTTCTTGATGTA 61106

RESULT 11  
AC125486

LOCUS 281892 bp DNA linear HTG 27-JUN-2002  
DEFINITION Mus musculus chromosome UNK clone RP23-21G7, WORKING DRAFT  
SEQUENCE, 15 unordered pieces.

ACCESSION AC125486  
VERSION AC125486.1 GI:21617767  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 281892)  
AUTHORS McPherson, J.D. and Waterston, R.H.  
TITLE The sequence of Mus musculus clone  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 281892)  
AUTHORS McPherson, J.D. and Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Project Information -----  
Center project name: M\_BA0021G07

----- Summary Statistics -----  
Sequencing vector: M13; 0%  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 274975 bases at least Q40  
Consensus quality: 276824 bases at least Q30  
Consensus quality: 278008 bases at least Q20  
Insert size: 230000; agarose-fp  
Insert size: 280492; sum-of-contigs  
Quality coverage: 11.26 in Q20 bases; agarose-fp  
Quality coverage: 7.91 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1622: contig of 1622 bp in length  
\* 1623 1722: gap of unknown length  
\* 1723 4017: contig of 2295 bp in length  
\* 4018 4117: gap of unknown length



/tissue\_type="cDNA-collection"  
/clone\_lib="686 (synonym: hicc3). Vector pSport1\_Sfi; host  
DH10B; sites SfiIA + SfiIB"  
/dev\_stage="adult"  
polyA\_signal 2246..2251  
polyA\_site 2256  
BASE COUNT 670 a 462 c 468 g 732 t  
ORIGIN

Query Match 11.2%; Score 403.8; DB 9; Length 2332;  
Best Local Similarity 99.5%; Pred. No. 1.5e-94;  
Matches 405; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 337 GAGAAGCAGTTAGAGTGTGCTTAACACACTGCAATTAAGTGCAATGCTGCTGGAC 396  
|||||  
Db 1 GAGAAGCAGTTAGAGTGTGCTTAACACACTGCAATTAAGTGCAATGCTGCTGGAC 60  
QY 397 CGAGCCCTGCCGCCCAACATCACCTGAAGGAGCTGCCATCTCTTTATCCTCTCTTTTCAT 456  
|||||  
Db 61 CGAGCCCTGCCGCCCAACATCACCTGAAGGAGCTGCCATCTCTTTATCCTCTCTTTTCAT 120  
QY 457 TCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTT 516  
|||||  
Db 121 TCTGCCAGTGACATTTTCAATGTTGCAAAACCAAAACCCCTTCTACCAATGCTCAGTT 180  
QY 517 GTTGTGTTTTCACAGTACTAAGAGTGTGAAGATGCGCCACTCTGGACTGCTTAAAGGAAAT 576  
|||||  
Db 181 GTTGTGTTTTCACAGTACTAAGAGTGTGAAGATGCGCCACTCTGGACTGCTTAAAGGAAAT 240  
QY 577 AGCAGACAGACATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAGTG 636  
|||||  
Db 241 AGCAGACAGACATGAGGGGCTACTTGACACAGATAAAGAGTCCCTGGATTAGTG 300  
QY 637 CTAATGCAAGATTGGCTTTCCTGAGTGATTTCCACCAACATTCGAAGAAACAATCAA 696  
|||||  
Db 301 CTAATGCAAGATTGGCTTTCCTGAGTGATTTCCACCAACATTCGAAGAAACAATCAA 360  
QY 697 CTAACCAACAAATGCGCAAGAAATCTTCTCTTAAAGTCAAACTGTT 743  
|||||  
Db 361 CTAACCAACAAATGCGCAAGAAATCTTCTCTTAAAGTCAAACTGTT 407

RESULT 13  
AK001821  
LOCUS Homo sapiens cDNA FLJ10959 fis, clone PLACE1000562. PRI 01-AUG-2002  
DEFINITION AK001821  
ACCESSION AK001821  
VERSION AK001821.1 GI:7023328  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens placenta cDNA to mRNA, clone\_lib:PLACE1  
clone:PLACE1000562.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,  
Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hata,R., Takeuchi,K.,  
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,  
Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.  
NEDO human cDNA sequencing project

TITLE Unpublished  
JOURNAL 2 (bases 1 to 2076)  
REFERENCE Isogai,T. and Otsuki,T.  
AUTHORS Direct Submission  
JOURNAL Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing and clone selection;  
Helix Research Institute (supported by Japan Key Technology Center

etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.  
FEATURES  
source Location/Qualifiers  
1..2076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="PLACE1000562"  
/tissue\_type="placenta"  
/clone\_lib="PLACE1"  
/note="Cloning vector: pME18SFL3"  
11..409  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAA91926.1"  
/db\_xref="GI:7023328"  
/translation="MIRTNSHVVGQLDDIRKNPRFVCLNDINIDHNHKAQIVKAVL  
RDFYESMFIPIQFOLPREYRNLHMHLEQWRAVYRDKLKFTHCVLATLIMFTIFS  
FFAQILALKRKIFPRRIHKEASPNRIRV"  
BASE COUNT 588 a 409 c 400 g 679 t  
ORIGIN

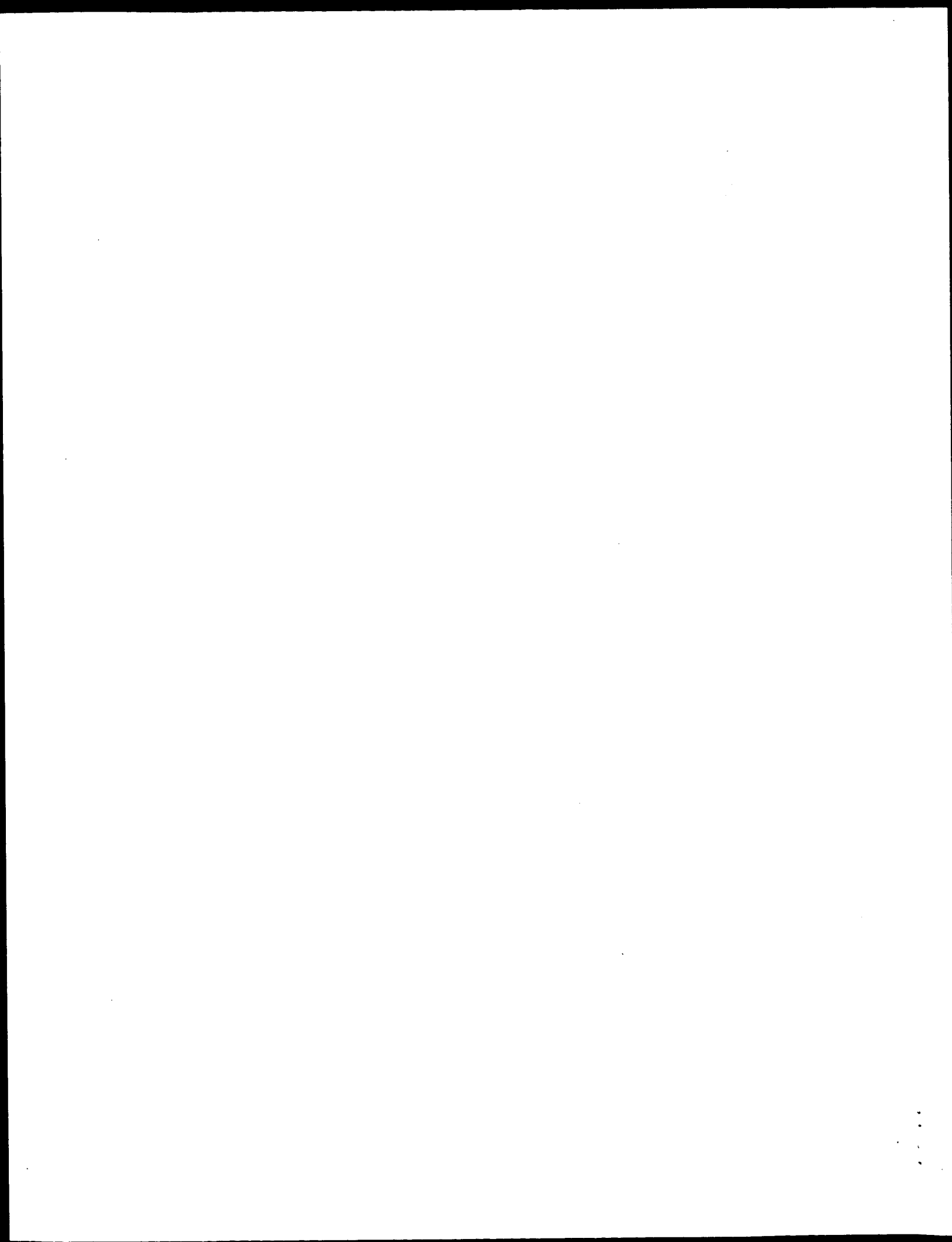
Query Match 7.4%; Score 266; DB 9; Length 2076;  
Best Local Similarity 100.0%; Pred. No. 2.1e-58;  
Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3333 CGCTTTTAAATGATTCCTACCAAGTTTCTCATGTGCTGCGCAGTTGGATGACATAAG 3392  
|||||  
Db 1 CGCTTTTAAATGATTCCTACCAAGTTTCTCATGTGCTGCGCAGTTGGATGACATAAG 60  
QY 3393 AAAAACCCCTAGGAAGTTTGTTCCTGAATGACACAACTTGACCACAACTATAAGATGC 3452  
|||||  
Db 61 AAAAACCCCTAGGAAGTTTGTTCCTGAATGACACAACTTGACCACAACTATAAGATGC 120  
QY 3453 TCAGACAGTGAAGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCTCCATACCTCCCA 3512  
|||||  
Db 121 TCAGACAGTGAAGCTGTTCTCAGGAGCTTCTATGAATCCATGTTCCTCCATACCTCCCA 180  
QY 3513 ATTTGAACCTGCAAGAGATATCGAAACCGTTTCTTCATATGATGAGCTGCAGGAATG 3572  
|||||  
Db 181 ATTTGAACCTGCAAGAGATATCGAAACCGTTTCTTCATATGATGAGCTGCAGGAATG 240  
QY 3573 GAGGCTTATCGACACAAATTTGAAGT 3598  
|||||  
Db 241 GAGGCTTATCGACACAAATTTGAAGT 266

RESULT 14  
AL135758  
LOCUS Mouse DNA sequence from clone Rp21-117J5 on chromosome 5, complete  
DEFINITION AL135758 142667 bp DNA linear ROD 29-JUN-2002  
sequence.  
ACCESSION AL135758  
VERSION AL135758.32 GI:9799912  
KEYWORDS HTG.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 142667)  
AUTHORS Blakey,S.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
clonerequests@sanger.ac.uk  
On Aug 13, 2000 this sequence version replaced gi:8978010.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:



QY 3061 CGAACACTGGCTACCGAATTTCAGAACTCCCGTTAAGTTTCAGGATTTCACAGGCTCG 3120  
Db 10160 CGAACACTGGCCACGAGAAATTCAGACCTACCTTTAAGCTTGCAAGTAAAGTATTGTTTT 10101  
QY 3121 GAACACATGCTAATAAATTG 3140  
Db 10100 AACCTTACGCTTGTGACGTG 10081

Search completed: April 13, 2003, 01:14:45  
Job time : 8201 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:20:41 ; Search time 29.9002 Seconds  
(without alignments)  
1179.859 Million cell updates/sec

Title: US-10-023-888-2  
Perfect score: 6340  
Sequence: 1 METDTLLWLLVLLVPGSTG.....NRELFHMHQLQWRYRDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	4.2	545	US-08-936-107A-10	Sequence 10, Appl
2	171.5	2.7	10182	US-09-134-001C-3159	Sequence 3159, Ap
3	166.5	2.6	1015	US-08-537-210A-1	Sequence 1, Appl
4	166.5	2.6	1015	US-09-113-825-1	Sequence 1, Appl
5	166.5	2.6	2471	US-08-185-432-16	Sequence 16, Appl
6	166.5	2.6	2471	US-08-083-590A-19	Sequence 19, Appl
7	166.5	2.6	2471	US-08-532-384-19	Sequence 19, Appl
8	166.5	2.6	2471	US-08-899-232-1	Sequence 1, Appl
9	163.5	2.6	3135	US-08-323-170B-2	Sequence 2, Appl
10	163.5	2.6	3135	US-08-954-441-2	Sequence 2, Appl
11	153	2.4	1964	US-09-467-997-1	Sequence 1, Appl
12	145.5	2.3	1068	US-08-537-210A-2	Sequence 2, Appl
13	145.5	2.3	1068	US-09-113-825-2	Sequence 2, Appl
14	145.5	2.3	2556	US-08-185-432-17	Sequence 17, Appl
15	145.5	2.3	2556	US-08-083-590A-20	Sequence 20, Appl
16	145.5	2.3	2556	US-08-532-384-20	Sequence 20, Appl
17	145.5	2.3	2556	US-08-899-232-2	Sequence 2, Appl
18	143.5	2.3	2285	US-09-308-375-2	Sequence 2, Appl
19	141	2.2	1078	US-08-264-534-32	Sequence 32, Appl
20	141	2.2	1078	US-08-083-590A-11	Sequence 11, Appl
21	141	2.2	1078	US-08-465-500-32	Sequence 32, Appl
22	141	2.2	1078	US-08-346-128-32	Sequence 32, Appl
23	141	2.2	1078	US-08-532-384-11	Sequence 11, Appl
24	141	2.2	1078	US-08-893-828-32	Sequence 32, Appl
25	139	2.2	219	US-08-902-516-2	Sequence 2, Appl
26	138	2.2	1664	US-08-537-210A-3	Sequence 3, Appl
27	138	2.2	1064	US-09-113-825-3	Sequence 3, Appl

28 138 2.2 2523 1 US-08-185-432-18 Sequence 18, Appl  
29 138 2.2 2523 4 US-08-899-232-3 Sequence 3, Appl  
30 137 2.2 1388 4 US-09-572-191-2 Sequence 2, Appl  
31 137 2.2 1388 4 US-09-723-262-2 Sequence 2, Appl  
32 137 2.2 1388 4 US-09-723-219-2 Sequence 2, Appl  
33 137 2.2 3696 4 US-09-134-001C-5080 Sequence 5080, Ap  
34 136 2.1 878 4 US-09-134-001C-4578 Sequence 4378, Ap  
35 134 2.1 1211 4 US-09-134-001C-4820 Sequence 4820, Ap  
36 134 2.1 1852 1 US-08-425-061-24 Sequence 24, Appl  
37 134 2.1 1852 1 US-08-825-886-24 Sequence 24, Appl  
38 134 2.1 1863 1 US-08-425-061-16 Sequence 16, Appl  
39 134 2.1 1863 1 US-08-480-784-2 Sequence 2, Appl  
40 134 2.1 1863 1 US-08-483-553-2 Sequence 2, Appl  
41 134 2.1 1863 1 US-08-487-002-2 Sequence 2, Appl  
42 134 2.1 1863 1 US-08-483-554B-2 Sequence 2, Appl  
43 134 2.1 1863 1 US-08-488-011B-2 Sequence 2, Appl  
44 134 2.1 1863 2 US-08-825-886-16 Sequence 16, Appl  
45 134 2.1 1863 4 US-08-850-727-2 Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-08-936-107A-10  
; Sequence 10, Application US/08936107A  
; Patent No. 6403306  
; GENERAL INFORMATION:  
; APPLICANT: Stephens, David S.  
; TITLE OF INVENTION: Serogroup-Specific Nucleotide Sequences  
; TITLE OF INVENTION: in the Molecular Typing of Bacterial Isolates and the  
; TITLE OF INVENTION: Preparation of Vaccines Thereo  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: US  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 23-SEP-1997  
; APPLICATION NUMBER: US/08/936,107A  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/827,622  
; FILING DATE: 09-APR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Caruthers, Jennie M.  
; REGISTRATION NUMBER: 34,464  
; REFERENCE/DOCKET NUMBER: 77-97  
; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 545 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-936-107A-10

Query Match 4.2%; Score 264; DB 4; Length 545;  
Best Local Similarity 27.8%; Pred. No. 2.5e-13;  
Matches 90; Conservative 52; Mismatches 130; Indels 52; Gaps 13;

QY 135 LPANITLDSLPSFFHSASDIFNVAKPNPSNVVVFDSTKDVEDHSLGK-NS 193

Db 97 LPSNLTLK--PALCIILESHKEDFLN-----KFLLTISSENULKIQKFNQIKNPKS 145  
Qy 194 ROTWRGY-----LTTDKKEVPGVL-----LMQDLAFLSGFPPTFKETNQLKT 235  
Db 146 VNEIWTDLFSIAHVDMKLSLSSISQFWFRLEFCREKDFI-----LFSTANRYSK 200  
Qy 236 KLPELSSKYLQLYSEASVALLKNNPKDFQELANKQTKKNTIDGKELTISPAYLLWD 295  
Db 201 KLWKH-----SIKNQLKEG-----IRNYSISSLPEEDHNFIDLVFTWVNSDEKNO 251  
Qy 296 LSATSQSKQD---EDISARFEDNEELRYSLSIERHAPWVRNFIPTVNGQIPSWNLND 352  
Db 252 -ELYKTKPDNFNSDATSFSLRDELFAIRSWEMSGFIRKIFIVSNCAPPAILDLNN 310  
Qy 353 PRVTIVHQDFVFNRLSHLTFSSPAIESHVHRIEGLSKOFIYLLNDVDFKQWMPDDFYS 412  
Db 311 PKIQWVHEEIMPO-SALPTFSSHAIEISLHHIPGISNFIYSNDDFLTKPLNKDNFFY 369  
Qy 413 HSKQKQVYL-TWVPVN-CAGGCP 433  
Db 370 SNGIAKRLLEAWGNVNGECTEGEP 393

RESULT 2  
US-09-134-001C-3159  
; Sequence 3159, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134, 001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064, 964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055, 779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3159  
; LENGTH: 10182  
; TYPE: PRN  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3159

Query Match 2.78; Score 171.5; DB 4; Length 10182;  
Best Local Similarity 17.28; Pred. No. 0.0015;  
Matches 221; Conservative 212; Mismatches 461; Indels 389; Gaps 57;

Qy 15 VPGSTGDEQDPRIDGKLSRDYH---VLFDSYRONIAGKSPQNRCL-PMPIDVVT 70  
Db 7255 VPSSTLDRATIDALLEQNARESLHGEOKLOEAKNAVA--EIDNLOALNPGVLAET 7312  
Qy 71 WYN---GTDLELLKELQVREOEEQAMREILGKNTTEPTKSEKOLECLLTHCIKVP 128  
Db 7313 LVNQASTPEVQEQALQAKE-LNEAMKALK-----TEINKKEQIKADSRVNA----- 7359  
Qy 129 LVIDPALPANITKLDPLSLYFSFHSASDFIVNAKPKNPSTNVSVVFDSTKVEDAHSL 188  
Db 7360 ---DSGLQANYN-----SALNYGSOIATQTP--PELNKD-VINRATQTIKTAENN- 7404  
Qy 189 LKGNRSQTWVRGYLTDTTKEVPLVMQDLAFLSGFPPTFKE-----TNQLKTK-----LP 238  
Db 7405 LNGQSK-----LAEAKSDG---NQSIEHLQGLTOSQKDKQHDLDLNOAQTKQOQVDIV 7453  
Qy 239 ENL-----SSKVKLLQLYSEASVALLKNNPKDFQELNKQTKKNTIDGKELTISPAYLLW 294  
Db 7454 NNSKQLDNSMNOLOQIVNNDNT--VKQNSDFINEDSSQODAYNHAIOAAKDLITAHPTIM 7511  
Qy 295 DLSAISQSKQD-----EDISAR--FEDNEELRYSLSIERHAPWVRNFIPTVNGQIPSW 347  
Db 7512 DKNQIDQAIENIKALNDLHGSNKLSEDKKASEOL-----QMLNSLTNGQ----- 7557

Qy 348 LNLNPRVTIVTHQDFVFNRLSHLTFESSA---IESHVHRIEGLSKOFIYLLNDV--- 399  
Db 7558 -----KDTILNH-----IFSAPTSQYGEKIASAKQLNNTMKALRDSIADNE 7600  
Qy 400 -----MEGKDVWPDDEFYSH--SRGQKQVYLTWVPNCAEGCPGSGWIKDGYCDKACNNSAC 451  
Db 7601 ILOSSKYFNEDESEQONAYNAQAVNAKNIINDQPTPVMANDEIQSVLNEVKTK----- 7653  
Qy 452 DWGSDCGSGNGSGSRYIAGGGGTGSGVGPQWGGGNGSINSYNOGNCANSWLADKFCDO 511  
Db 7654 -----DNLHGQKLANDKTDQA-----ATLNALNYLNQA-----Q 7683  
Qy 512 AGNVLSCFGDAGDCQDHFHLYKVIPLNQTHYIIPKGECLPYFSEFAEVAKRGVGAYS 571  
Db 7684 RGNL-----ETKQVNSNSRPEQKVQVQVLANQLNDAMK-----LDDALT 7722  
Qy 572 DNPIIRHASIAKWKTIHLIMHSGMNATTIHNLTFQNTDNEEFKMQITVEVDTRGPKL 631  
Db 7723 GNDALIKQTS-----NYINEDTSQQVNFDEYDRG--- 7751  
Qy 632 NSTAQGYENLVSPITLLPEAEILFEDIPKRRPKFKRRHDVNSTRRRAQAEVKIPLVNIS 691  
Db 7752 -----KNIVA-----EOTNPMGPTNINTIADKITAKNDLHGVO 7786  
Qy 692 LLPKDAQLSLNTLDLQLEHGDTTLKGYNLSKLSALLRSFLMSQ-----HAKIKNQAIITD 746  
Db 7787 KLKQAOQOSINTIN-----QMTGLNQAOKEQLNQEIQTQTRSEVHVQVINKAQAALND 7838  
Qy 747 ETN---DSLVAPQEQKVHKSILPNSLG-----VSELRQLRITFPVAVSKVNGHDGONP 796  
Db 7839 SMNTLRQSTIDHEVQKTSNYINETVGNQATAYNNAVDKVKQII-----NQTNP 7887  
Qy 797 ---PLDLE--TTAREVE---THQKTIGNTVKKEKPPSLIPLVESQMTKEKKITGKEKEN 849  
Db 7888 TWNPLEVERATSNVKISKDALHGERELNDN--KNSKTTAVNHLNDLNAQAKEALTHEIEQ 7945  
Qy 850 SRMEENAENHIGVTEVLIG--RKLOHYTDSYLGFLPWEKKYKFFLDLDBEESLKTQLAYF 907  
Db 7946 ATIVSQVNNIYNKAKALNNDMKLK-----DIVAQDNVNRQSNNYI 7986  
Qy 908 TDSKNRARKYKRTDFADSLRYVNVKILNSKEGFTSRKVPAPHMHIDRIVMQELQDMFPEEF 967  
Db 7987 NEDSTPQNNYNDTINHAQSIIDQVANPT-----MSH--DEI----- 8020  
Qy 968 DKTSFKVHRHS-----EDMQFAFSFYFLMSAVQPLN-----ISOVFEVDVTDQSGV 1014  
Db 8021 -ENAINNIKHAINALDGEHKLQOAKENANLLNSLNDLNAPORDAINRLVNEAQTRKVA 8079  
Qy 1015 LSDREIRTLATRIHELPLSLQDLTGL--EHLIINGSKMLPADITQINLIPTQTESYYPDN 1072  
Db 8080 EOLOSQAALNDAMKHLRNSIQNSSVRQESKYINAS---DARKEQYNHAREVENIINEQ 8136  
Qy 1073 LPVYTKSLVNTCKPVTDKTHKAYKDKNKYRFEIMG-----E 1108  
Db 8137 HPTLDKEII---KQLFDGVNQANNLDN--GVLELDAQKONAHQSIPTLHMLNQAQONALN 8191  
Qy 1109 EETAFKMITNVSHVVGQ---LDDIRKNPRKEVCLNDNIDHNHDKAQTVKAVLRDPFYSM 1165  
Db 8192 EKINNATVTEVAALIGQAKLLDHAM-----ENLEESIKDKEQVK-----QSS 8234  
Qy 1166 FPIPSQFELPREYRNFRLMHHEL 1188  
Db 8235 NYINEDSDVOETYNDAVDHVTET 8257

RESULT 3  
US-08-537-210A-1  
; Sequence 1, Application US/08537210A  
; Patent No. 5780300  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Fortini, Mark



TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2471 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-185-432-16

Query Match 2.6%; Score 166.5; DB 1; Length 2471;  
Best Local Similarity 19.7%; Pred. No. 0.00039;  
Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;

QY 424 PYPNCAEGCPGSGWIKGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGISGVGPW 483  
DB 1421 PPATCLSQYCADKARDGVCDCEACNSHACQWGGDC-----SLTMENPW 1463  
QY 484 -QFGGINSVSYCNGC-----ANS-----WLDKF-----CDQACN 514  
DB 1464 ANCSPLPCWDYINNQCDELNTVECLFDNFECQNSKTKYDKYCADHFKDNHNOGCN 1523  
QY 515 VLSGFDAGDCQDHFHLYK-----VILLPNQTHYIIPKGECLPYFSAEV-----561  
DB 1524 SEECWMDGLDCAADQPENLAEGTLVIVLMP-----PEQLQDARSFLRALGTLHTN 1576  
QY 562 --AKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFQNTDEEFK---616  
DB 1577 LRIKRDQSGLMVPYGYGKSAAMKKQ-----RMTRSLPGEQEVEA 1619  
QY 617 -MQITVEVDTRGPKLNSTAQGYENLVSPITLLPEAEILFEDIPEKRFKFRHDVNS 675  
DB 1620 GSKVFEIDNRQ---CVQDSHDFKNTDAAALAS-----HAIOG 1657  
QY 676 TRRAQEEVKIPLVNI---SLLPKDAQSLNTLDLQL-----EHGDI 714  
DB 1658 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVIILFILLGVIMAKRKHGSLW 1710  
QY 715 L-KGYNLKSALLRSLFNMNSOHAKE-----IKNOAITDETNDLSIVAPQEKOVHKS 763  
DB 1711 LPEGFTLRDA-----SNHRRPVGQDAVGLKNLSVQVSEAN--LI-----1750  
QY 764 ILPNSLGVSERLQRLTFPAVSVKVNHDGQ-----QNPPLDLETTARFRVETHTQKTIGN 819  
DB 1751 ---GTGTSEHWVDDEGPQPK-KVKAEDAEALLSEEDDPIDR-----RPTQOHLLEA 1797  
QY 820 VTKEKPPSLIVPLESOMTKE-----KKITGKEKENSMEENAEH---H 859  
DB 1798 DIRTPSLALTPQAEQEVLDVNVVRGPDGCTPLMLASLRGSSDLSDEDEDAEDSSAN 1857  
QY 860 IGVTEVLLGRKLQHYTD 876  
DB 1858 IITDLVYQGASLQAQTD 1874

RESULT 6  
US-08-083-590A-19  
Sequence 19, Application US/08083590A  
Patent No. 5786158  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, S. et al.  
TITLE OF INVENTION: Therapeutic And Diagnostic Methods  
TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And  
TITLE OF INVENTION: Nucleic Acids  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:

QY 484 -QFGGINSVSYCNGC-----ANS-----WLDKF-----CDQACN 514  
DB 310 ANCSPLPCWDYINNQCDELNTVECLFDNFECQNSKTKYDKYCADHFKDNHNOGCN 369  
QY 515 VLSGFDAGDCQDHFHLYK-----VILLPNQTHYIIPKGECLPYFSAEV-----561  
DB 370 SEECWMDGLDCAADQPENLAEGTLVIVLMP-----PEQLQDARSFLRALGTLHTN 422  
QY 562 --AKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFQNTDEEFK---616  
DB 423 LRIKRDQSGLMVPYGYGKSAAMKKQ-----RMTRSLPGEQEVEA 465  
QY 617 -MQITVEVDTRGPKLNSTAQGYENLVSPITLLPEAEILFEDIPEKRFKFRHDVNS 675  
DB 466 GSKVFEIDNRQ---CVQDSHDFKNTDAAALAS-----HAIOG 503  
QY 676 TRRAQEEVKIPLVNI---SLLPKDAQSLNTLDLQL-----EHGDI 714  
DB 504 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVIILFILLGVIMAKRKHGSLW 556  
QY 715 L-KGYNLKSALLRSLFNMNSOHAKE-----IKNOAITDETNDLSIVAPQEKOVHKS 763  
DB 557 LPEGFTLRDA-----SNHRRPVGQDAVGLKNLSVQVSEAN--LI-----596  
QY 764 ILPNSLGVSERLQRLTFPAVSVKVNHDGQ-----QNPPLDLETTARFRVETHTQKTIGN 819  
DB 597 ---GTGTSEHWVDDEGPQPK-KVKAEDAEALLSEEDDPIDR-----RPTQOHLLEA 643  
QY 820 VTKEKPPSLIVPLESOMTKE-----KKITGKEKENSMEENAEH---H 859  
DB 644 DIRTPSLALTPQAEQEVLDVNVVRGPDGCTPLMLASLRGSSDLSDEDEDAEDSSAN 703  
QY 860 IGVTEVLLGRKLQHYTD 876  
DB 704 IITDLVYQGASLQAQTD 720

US-08-185-432-16  
Sequence 16, Application US/08185432  
Patent No. 5750652  
GENERAL INFORMATION:  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Busseau, Isabelle  
APPLICANT: Diederich, Robert J.  
APPLICANT: Xu, Tian  
APPLICANT: Matsuno, Kenji  
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND  
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/185,432  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090



Db 1711 LPEGFTLRDA-----SNHKREPVGQDAVGLKNLSVQVSEAN--LI-----1750  
QY 764 ILPNSLGSVERLQRLTFPAVSKVNGHDG-----QNPPLDLETTARFRVETHQKTIGGN 819  
Db 1751 ----GTGISEHWVDDEGPQPK-KVKAEDALLSEDDPIDR-----RPWTQOQHLEAA 1797  
QY 820 VTKEKPPSLIVPLESOMTKE-----KKITGKEKENSMEENAE--H 859  
Db 1798 DIRRTPSLALTPPQAEQEVVDVNVVRPGDCTPLMLASLRGGSSDLSEDEDAEDSSAN 1857  
QY 860 IGVTVELLGRKLQHYTD 876  
Db 1858 IITDLVYQGASLQAQTD 1874  
RESULT 8  
US-08-899-232-1  
; Sequence 1, Application US/08899232  
; Patent No. 6436650  
; GENERAL INFORMATION:  
; APPLICANT: Artavanis-Tsakonas, Spyridon  
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON  
; FILE REFERENCE: 7326-046  
; CURRENT APPLICATION NUMBER: US/08/899,232  
; CURRENT FILING DATE: 1997-07-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2471  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-08-899-232-1

Query Match 2.6%; Score 166.5; DB 4; Length 2471;  
Best Local Similarity 19.7%; Pred. No. 0.00039;  
Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;  
QY 424 PVPNCAECPCGSKWIKGDCYCDKACNNSACDWDGDCSGNSGSRVYAGGGTSGVGPW 483  
Db 1421 PRATCLSOYCADKADGVCDEACNSHACQWDGDC-----SLTMENPW 1463  
QY 484 -QFGGINSVSYCNQGC-----ANS-----WLADKF-----CDQACN 514  
Db 1464 ANCSSPLFCWDYINNQCDELNCTVECLDFNFCQGNSTCKYDKYCADHFKDNCNCGN 1523  
QY 515 VLSGCFDAGDCGQHFHLYK-----VILLPNQTHYIIPKGECLPYFSFAV-----561  
Db 1524 SEECGWDGLDCAADQENLAEGTLIVVILMP-----PEOLLQDARSFLRALGTLHTN 1576  
QY 562 --AKRGVEGAYSDNPIIRHASTANKWKTIIHMHSGMNATTIHFNLTFONTNDEEFK---616  
Db 1577 LRIKDSOGELMVYPYCEKAAMKQ-----RMTFRSLPGEQGEVA 1619  
QY 617 -MQITVEDTREGPKLNSTAGQENLVSPITLLPEAEILFEDIPKEKRPFKRHDVNS 675  
Db 1620 GSKVFEIDNRQ---CVQDSHDCFKNTPDAAAALLAS-----HAIQG 1657  
QY 676 TTRAEVVKIPLVNI---SLLPKDAQLSLNTLDLQL-----EHGDI 714  
Db 1658 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVIILFIILGVIMAKRKRKHSGLW 1710  
QY 715 L-KGYNLKSALLRSFLMNSQAK-----IKNQAITDETNTSLVAPQEKQVHKS 763  
Db 1711 LPEGFTLRDA-----SNHKREPVGQDAVGLKNLSVQVSEAN--LI-----1750  
QY 764 ILPNSLGSVERLQRLTFPAVSKVNGHDG-----QNPPLDLETTARFRVETHQKTIGGN 819  
Db 1751 ----GTGISEHWVDDEGPQPK-KVKAEDALLSEDDPIDR-----RPWTQOQHLEAA 1797  
QY 820 VTKEKPPSLIVPLESOMTKE-----KKITGKEKENSMEENAE--H 859  
Db 1798 DIRRTPSLALTPPQAEQEVVDVNVVRPGDCTPLMLASLRGGSSDLSEDEDAEDSSAN 1857

QY 860 IGVTVELLGRKLQHYTD 876  
Db 1858 IITDLVYQGASLQAQTD 1874

RESULT 9  
US-08-323-170B-2  
; Sequence 2, Application US/08323170B  
; Patent No. 5733772  
; GENERAL INFORMATION:  
; APPLICANT: Williamson, Kim C.  
; APPLICANT: Kaslow, David C.  
; TITLE OF INVENTION: Cloning and Expression of Plasmodium  
; TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pfs230  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/323,170B  
; FILING DATE: 13-OCT-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/010,409  
; FILING DATE: 29-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Quine, Jonathan A.  
; REGISTRATION NUMBER: P-41,261  
; REFERENCE/DOCKET NUMBER: 015280-113100US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3135 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-323-170B-2

Query Match 2.6%; Score 163.5; DB 1; Length 3135;  
Best Local Similarity 19.7%; Pred. No. 0.001;  
Matches 206; Conservative 124; Mismatches 378; Indels 337; Gaps 52;

QY 236 KLPLENSSKVKLLQLYS-EASVALLKLNPK-----DFQLN-----KQTKK 276  
Db 1654 QLEEVYNDIESLELDIEQYVLQVNLKAPKLMMSAQIHNNRHVCDFSKNLIVPELKK 1713  
QY 277 NMTIDCKELTIPAYLLWDLISAI-----SOSKODEDISASRFEDNE---ELR-----320  
Db 1714 KEELGNGPNVNIHCYALLKPLDLYVKCPTSKDNEYAAKNISENDNEYELQVLSIEKRF 1773  
QY 321 YLSRISIERHAPVVRNFIIVTNGQIPSWLALNDPRVTIVTHQDVRNLSHLPTFSSPAIS 380  
Db 1774 HNFETLESKKPGNGDV-VVHNGVVDTPGLDNS-----TFEYFKNIKIKP-----1818  
QY 381 HVRIEGLSKFIYLLNDVWFQKD---VVPDDFEYSHSKGQK-----VLTWVPVPCARG 431  
Db 1819 -----DKFFEKVINEYDTEEEKDLESILPGAIVSPMKVILKKKDPPTSTAAAFVVPPIV- 1872  
QY 432 CPGSWTIKDGYCDKACNNSACDWDGDCSGNSGSRVYAGGGTSGVGPQ-----WQF 485  
Db 1873 -----KDLHFKECNTE-----YKDENOYISYNGIIHIDISNKRKINGCDF 1916

```

QY 486 GGGINSV-----SYCNQGCANSLADKFCDOACNVLSGCFDAGCGQDHFHEL 533
Db 1917 STNNSILTSVSKVLNGETKNCENINNNVEFGIICDNETNL-----DPEKC---FHEI 1967
QY 534 YKVIILLPNOT---HYIIPKGECLPYFSAEVAKRGEVAGSDNPI--IRHASIANKWT 587
Db 1968 YS---KDNKTVKKFEVIPN---IDIFSLHNSNKKV--AYAKVPLDYINKLLFSCSCKT 2019
QY 588 IHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTRREGPKLNSTAQKGYENLVSPIT 647
Db 2020 SHNTIGTKMVT---LNKDEKEEDFK-----TAQGIKHNHVHLCN 2057
QY 648 LLPEAEILFEDIPKEKRPFKRHDVNSTRRRAQEEVKIPLVNIISLLPKDAQLSLNTLDLQ 707
Db 2058 FFDNPELTFDN-----NKIVLCKI-----DAEL-FSEVIQ 2087
QY 708 L-----EHGDIITLKGYNLSKSALLRSLMNSOHAK--IKNOAIITDETNDLSVAPQE 757
Db 2088 LPFQTKNVEEG-----VONEEYKFKSLKPSLVFDDNNNDIKVIGKE 2129
QY 758 KQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGN-----PPLDLETTARFRV-ET- 810
Db 2130 K-----NEVSIISLAKGVYGNRIFTDKNGKKGEGISFFIPIKODTDLKFIIINETI 2181
QY 811 -----HTQKTIGGNVTK-----EKPPSLIVPLESQMTKEKKITGKEKENSME 853
Db 2182 DNSNIKQRLIIVFVRKNVSENSFKLCDFTTGSTSLMELNSQV-KEKCTVKKIKKGDFG 2240
QY 854 ENAENHIGV-----TEVLLGRKLOHYTDS-YLGFLPWEKKY-----FLDLDEESL 900
Db 2241 LKCPKGFATFPQACFSNVLLEYKSDYEDSEHINYIHKDKKYNLKPDKVIELMDEN--- 2297
QY 901 KTQLAYFDSKNRARKYR-DFFADSLRYVN-KILNSREGFTSRKVPAR--MPHMIDRIVM 956
Db 2298 -----FRELNIQOYTGISNITDVLHFKNENLGNLPLNFKNHSTAYAKVPDFTFSIIN 2351
QY 957 QELQDMFPE-----EFDKTSFHKVHSEDMQFAF-----SYFYVILMSAVQPLNIS 1001
Db 2352 FSCCYNPEKHYVTGMQVESDNRFNDIKKNENYIKNFLPNIEKYALLLDDEERQKKIK 2411
QY 1002 QVDFEVDVDQSVLSDREIRLATRIHELPLSLQDLTGLEHMLI----- 1045
Db 2412 QOQEEQOQEQ--ILKQDDR-----LSRHDDYKNKHTYILYDSNEHICDYEKNES 2459
QY 1046 -----NGSKMLPADITQLN-----NIPPTQESYDNPPLPVTKSLVTNC 1084
Db 2460 LISTLPNDTKKIQKSICKINAKALDVVTIKCPHTKNTPKD---YFFN-----SSLTND 2511
QY 1085 KPVTDKIHK-----AYKDKNKYRFEI 1105
Db 2512 KKVITFDKKNFVYIDPTKKTFSL 2536

```

## RESULT 10

```

US-08-954-441-2
; Sequence 2, Application US/08954441
; Patent No. 6316000
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,441
; FILING DATE: 20-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/323,170
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 015280-1131100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-954-441-2

```

```

Query Match 2.6%; Score 163.5; DB 4; Length 3135;
Best Local Similarity 19.7%; Pred. No. 0.001;
Matches 206; Conservative 124; Mismatches 378; Indels 337; Gaps 52;

```

```

QY 236 KLPENSSKVKLLQLYS-EASVALLKLNPK-----DFELN-----KQTKK 276
Db 1654 QLEEVYNDIESLELKDIEQYVLQVNLKAPKLMMSAQIHNRRVCDFSKNLLIVPESLKK 1713
QY 277 NMTIDCKELTISPAYLLWLSAI-----SOSKODEDISASRFEDNE---ELR----- 320
Db 1714 KEELGSPVNIHCYALLKPLDLYVKCPTSKONYEAKYINSENDNEYELQISLIEKRF 1773
QY 321 YSLRSTERHAPWVRNIFVINGQIPSWLNLDNPRVITVHQDVFRNLHLPTFSSPAIES 380
Db 1774 HNPETLESKKPGNDV-VVHNGVVDTPGVLDNS-----TFEYFKNIKIKP----- 1818
QY 381 HVHRIEGLSQKFIYLDNDVYMGKD---VMPDDFYSKSKQK-----VLTWPVPCABG 431
Db 1819 -----DKFEKVINEXDDTTEEEKDLESILFGAIVSPMKVLKKKDPFTSYAAFWVPIVP- 1872
QY 432 CPGSWIKDGYCDKACNNSACDWDGDCSGNSGSRVYAGGCTGSGVGP-----WQF 485
Db 1873 -----KDLHFKVECNTE-----YKDENVYISYNGIIHIDISNKRKINGCDF 1916
QY 486 GGGINSV-----SYCNQGCANSLADKFCDOACNVLSGCFDAGCGQDHFHEL 533
Db 1917 STNNSILTSVSKVLNGETKNCENINNNVEFGIICDNETNL-----DPEKC---FHEI 1967
QY 534 YKVIILLPNOT---HYIIPKGECLPYFSAEVAKRGEVAGSDNPI--IRHASIANKWT 587
Db 1968 YS---KDNKTVKKFEVIPN---IDIFSLHNSNKKV--AYAKVPLDYINKLLFSCSCKT 2019
QY 588 IHLIMHSGMNATTIHFNLTFQNTNDEEFKMQITVEVDTRREGPKLNSTAQKGYENLVSPIT 647
Db 2020 SHNTIGTKMVT---LNKDEKEEDFK-----TAQGIKHNHVHLCN 2057
QY 648 LLPEAEILFEDIPKEKRPFKRHDVNSTRRRAQEEVKIPLVNIISLLPKDAQLSLNTLDLQ 707
Db 2058 FFDNPELTFDN-----NKIVLCKI-----DAEL-FSEVIQ 2087
QY 708 L-----EHGDIITLKGYNLSKSALLRSLMNSOHAK--IKNOAIITDETNDLSVAPQE 757
Db 2088 LPFQTKNVEEG-----VONEEYKFKSLKPSLVFDDNNNDIKVIGKE 2129
QY 758 KQVHKSILPNSLGVSERLQRLTFPAVSVKVNGHDQGN-----PPLDLETTARFRV-ET- 810

```

Db 2130 K-----NEVSISLAKGVYGNRIFFDKNGKKGEGISFFPIPKQDTDLKFIINEI 2181  
Qy 811 -----HTQKTIIGNVTK-----BKPPSLVPLESQMTKTKKTKGKEKNSRME 853  
Db 2182 DNSNIQKORGLIYIFVRKNVSENSEFKLDCFTTGTSTSLMELNSQV-KEBKCTVKIKKGDIFG 2240  
Qy 854 ENAENHIGV-----TEVLGRKLOHVTDS-YLGFPLPWEKKKY-----FLDLLDEESL 900  
Db 2241 LKCPKFAIFPQAFCSNVLLYKSYDESEHINYIHKDKYNLKPDKVDIELMDEN---- 2297  
Qy 901 KTOLAYFTDSKNRARKR-DTFADSLRYVN-KILNSKGFSTSRKVPAAH-MPHMIDRIYM 956  
Db 2298 -----FRELQNIQYTGISNITDVLHFKNFNLGNLPLNFKNHYSTAYAKVPDTFNSLI 2351  
Qy 957 QELQDMPPE-----EFDKTSFKVHRSEDMQAF-----SYFYILMSAVQPLNIS 1001  
Db 2352 FSCNCPNPKHVYGTQMVESDNRFNFIKKNVNIKFNLLPNIEKYALLDLDEERQKKIK 2411  
Qy 1002 QVFEVDVTDGSGVLSDRIRTLATRIHELPLSLQDLTGLEHMLI----- 1045  
Db 2412 QQEEEOQEQ-----ILKQDDR-----LSRHDDYKNKNTVILYDSNEHICDYKNES 2459  
Qy 1046 -----NCSKMLPADITQLN-----NIPPTQESYDPLNPPVTKSLVTNC 1084  
Db 2460 LISTLPNDTKKIQKSICKINAKALDVVTIKCPHTKNFTPKD---YFPN-----SSLITND 2511  
Qy 1085 KPVTDKTHK-----AYKDKNKYRFEI 1105  
Db 2512 KKIIVTFDKKNFVYIIDPTKKTFSL 2536

## RESULT 11

US-09-467-997-1  
; Sequence 1, Application US/09467997  
; Patent No. 6379925

; GENERAL INFORMATION:

; APPLICANT: Kitajewski, Jan

; APPLICANT: Uytendaele, Hendrik

; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION

; FILE REFERENCE: 53863-A-PCT-US

; CURRENT APPLICATION NUMBER: US/09/467,997

; CURRENT FILING DATE: 1999-12-20

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 1964

; TYPE: PRT

; ORGANISM: mouse

US-09-467-997-1

Query Match 2.4%; Score 153; DB 4; Length 1964;  
Best Local Similarity 28.1%; Pred. No. 0.0036;  
Matches 41; Conservative 11; Mismatches 32; Indels 62; Gaps 8;

Qy 424 PVPNC-----AEGCPGSEWTKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGSIGV 479

Db 1162 GPQRCPRGASGCEBGRG-GDGTCDAGCGSGPGGDWDGDC-----SLGV 1203

Qy 480 GQWQFGGGINSVYC-----NQCA-----NSWLADKF- 508

Db 1204 PDPWK---GCPPHSOWLLFRDGRCHQCDSECLFDGYDCEIPTTCIPAYDQICRDFH 1260

Qy 509 ---CDQACNVLSGCFDAGDC---GQD 528

Db 1261 NGHCEKGCNNAECGWDGDCRPEGED 1286

## RESULT 12

US-08-537-210A-2

; Sequence 2, Application US/08537210A

; Patent No. 5780300

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon  
; APPLICANT: Fortini, Mark  
; APPLICANT: Matsuno, Kenji  
; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,210A  
; FILING DATE: 29-SEP-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 7326-027  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1068 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Human N1 (TAN-1)  
; LOCATION: 1152...2219  
; OTHER INFORMATION: Highly conserved ankyrin repeat  
; OTHER INFORMATION: region of No. 5780300ch

US-08-537-210A-2

Query Match 2.3%; Score 145.5; DB 1; Length 1068;  
Best Local Similarity 23.7%; Pred. No. 0.0055;  
Matches 50; Conservative 13; Mismatches 57; Indels 91; Gaps 11;

Qy 425 VFNCAEGCPGSEWTKDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGSIGVGPWQ 484

Db 301 LPQCQEDAGNK-----VCSLQCNHACGWDGDC-----SLNFDNPK 338

Qy 485 -----FGGINSVSYCNQ-GC-----ANSLADKFCDO 511

Db 339 NCTQSLOQWKYFSDG-HCDSQCSAGCLFDGFCQRAEGCQCNPLYDQYCKDHFSDGHCDO 397

Qy 512 ACNVLSCGFADGCGODHFE-----LYKVILLP-----NOTHYIIPKGECL----- 553

Db 398 GCSNABCEWDGLDCA-EHVPERLAAGTLVVVVVIMPEQLRNSFSFELMELSVLHTNVVF 456

Qy 554 -----PYESFAE-----VAKRGVEG 568

Db 457 KRDAHQQMIFPYGREGELRKHPIKRAAEG 487

## RESULT 13

US-09-113-825-2

; Sequence 2, Application US/09113825

; Patent No. 6149902

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Fortini, Mark

; APPLICANT: Matsuno, Kenji



;; TITLE OF INVENTION: MANIPULATION OF NON-TERMINALLY  
;; TITLE OF INVENTION: DIFFERENTIATED CELLS USING THE NOTCH PATHWAY

;; NUMBER OF SEQUENCES: 4

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Pennie & Edmonds

;; STREET: 1155 Avenue of the Americas

;; CITY: New York

;; STATE: NY

;; COUNTRY: USA

;; ZIP: 10036/2711

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible

;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq Version 2.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/113,825

;; FILING DATE:

;; CLASSIFICATION:

;; APPLICATION NUMBER: 08/537,210

;; FILING DATE: 29-SEP-1995

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Mistrock, S. Leslie

;; REGISTRATION NUMBER: 18,872

;; REFERENCE/DOCKET NUMBER: 7326-027

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 212-790-9090

;; TELEFAX: 212-869-8864

;; TELEX: 66141 PENNIE

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 1068 amino acids

;; TYPE: amino acid

;; STRANDEDNESS:

;; TOPOLOGY: unknown

;; MOLECULE TYPE: protein

;; FEATURE:

;; NAME/KEY: Human N1 (TAN-1)

;; LOCATION: 1152...2219

;; OTHER INFORMATION: Highly conserved ankyrin repeat

;; OTHER INFORMATION: region of No. 6149902ch

US-09-113-825-2

Query Match 2.3%; Score 145.5; DB 4; Length 1068;

Best Local Similarity 23.7%; Pred. No. 0.0055;

Matches 50; Conservative 13; Mismatches 57; Indels 91; Gaps 11;

QY 425 VPNAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGNSGSGRYIAGGGGTGSGIGVGPWQ 484

Db 301 LPEQEDAGNK-----VCSLQCNHACGWDGDC-----SLNFPDPMK 338

QY 485 -----FGGINSVSYCNO-GC-----ANSLADKFCDDQ 511

Db 339 NCTQSLQCNWKFSDG-HCDSQCNACGLFDGDCQRAEGCNPPLYDQYCKDHFSDGHCDDQ 397

QY 512 ACNVLSCGFDAGCQDHFHE-----LYKVILLP-----NQTHYIIPKGECL-----553

Db 398 GCNSAECWMDGLDCA-EHVPERLAAGTLVVVLMPPPEQLRNSFFHLWELSRVLTNNVVF 456

QY 554 -----PYFSFAE-----VAKRGVEG 568

Db 457 KRDAHQOQMIFPYGEEELRKHPIKRAAEG 487

RESULT 14

US-08-185-432-17

; Sequence 17, Application US/08185432

; Patent No. 5750652

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, Spyridon

; APPLICANT: Busseau, Isabelle

; APPLICANT: Diederich, Robert J.

;; APPLICANT: Xu, Tian

;; APPLICANT: Matsuno, Kenji

;; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

;; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

;; NUMBER OF SEQUENCES: 23

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: PENNIE & EDMONDS

;; STREET: 1155 Avenue of the Americas

;; CITY: New York

;; STATE: New York

;; COUNTRY: U.S.A.

;; ZIP: 10036-2711

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/185,432

;; FILING DATE: 21-JAN-1994

;; CLASSIFICATION: 530

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Mistrock, S. Leslie

;; REGISTRATION NUMBER: 18,872

;; REFERENCE/DOCKET NUMBER: 7326-006

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (212) 790-9090

;; TELEFAX: (212) 869-8864/9741

;; TELEX: 66141 PENNIE

;; INFORMATION FOR SEQ ID NO: 17:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 2556 amino acids

;; TYPE: amino acid

;; TOPOLOGY: unknown

;; MOLECULE TYPE: protein

US-08-185-432-17

Query Match 2.3%; Score 145.5; DB 1; Length 2556;

Best Local Similarity 23.7%; Pred. No. 0.023;

Matches 50; Conservative 13; Mismatches 57; Indels 91; Gaps 11;

QY 425 VPNAEGCPGSGWIKDGYCDKACNNSACDWDGDCSGNSGSGRYIAGGGGTGSGIGVGPWQ 484

Db 1452 LPEQEDAGNK-----VCSLQCNHACGWDGDC-----SLNFPDPMK 1489

QY 485 -----FGGINSVSYCNO-GC-----ANSLADKFCDDQ 511

Db 1490 NCTQSLQCNWKFSDG-HCDSQCNACGLFDGDCQRAEGCNPPLYDQYCKDHFSDGHCDDQ 1548

QY 512 ACNVLSCGFDAGCQDHFHE-----LYKVILLP-----NQTHYIIPKGECL-----553

Db 1549 GCNSAECWMDGLDCA-EHVPERLAAGTLVVVLMPPPEQLRNSFFHLWELSRVLTNNVVF 1607

QY 554 -----PYFSFAE-----VAKRGVEG 568

Db 1608 KRDAHQOQMIFPYGEEELRKHPIKRAAEG 1638

RESULT 15

US-08-083-590A-20

; Sequence 20, Application US/08083590A

; Patent No. 5786158

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

Search completed: April 13, 2003, 03:29:21  
Job time : 63.9002 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:34:46 ; Search time 66.7772 Seconds  
(without alignments)  
2392.545 Million cell updates/sec

Title: us-10-023-888-2

Perfect score: 6340

Sequence: 1 METDTLLVLLWVPGSTG.....NRFLLHMHLEQENRAYDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1893	29.9	367	22	Peptide #2930 enco
2	1893	29.9	367	22	Human brain expres
3	1893	29.9	367	22	Peptide #2891 enco
4	1893	29.9	367	22	Peptide #2989 enco
5	1893	29.9	367	22	Peptide #2868 enco
6	1893	29.9	367	23	Human peptide enco
7	854.5	13.5	652	22	Drosophila melanog
8	460	7.3	132	22	Human protein sequ
9	352	5.6	68	22	Peptide #5693 enco
10	352	5.6	68	22	Human brain expres

11	352	5.6	68	22	AAW71337
12	352	5.6	68	22	AAW71337
13	352	5.6	68	22	AAW71337
14	264	4.2	545	19	AAW79296
15	171.5	2.7	2469	23	AAE18207
16	171.5	2.7	10182	23	ABP38314
17	168.5	2.7	5024	22	AAW82935
18	166.5	2.6	2471	20	AAW06816
19	163.5	2.6	3135	15	AAW57474
20	163.5	2.6	3135	21	AAW18223
21	162.5	2.6	1654	6	AAW50777
22	162.5	2.6	1979	21	AAW18171
23	160	2.5	1963	22	AAW79838
24	157.5	2.5	1872	19	AAW68510
25	155.5	2.5	2321	19	AAW49698
26	154.5	2.4	1254	11	AAW07503
27	154.5	2.4	1254	18	AAW24575
28	153.5	2.4	1014	23	ABG61923
29	153.5	2.4	1960	22	AAW78854
30	153.5	2.4	2143	22	ABG01716
31	153	2.4	1964	20	AAW95557
32	152.5	2.4	7201	22	ABW71136
33	150.5	2.4	1639	19	AAW54145
34	150.5	2.4	2663	22	AAW39097
35	150.5	2.4	2688	22	AAW40883
36	148	2.3	2633	22	ABG06505
37	147.5	2.3	2954	20	AAW01632
38	147	2.3	3685	10	AAW90290
39	145.5	2.3	2444	23	ABW07821
40	145	2.3	28	22	AAW68261
41	145	2.3	962	22	AAW84956
42	143.5	2.3	2285	20	AAW98149
43	143	2.3	1078	13	AAW28963
44	143	2.3	6815	22	ABW66811
45	142	2.2	1109	22	AAU33434

#### ALIGNMENTS

#### RESULT 1

ABB30279

ID ABB30279 standard; Peptide; 367 AA.

XX ABB30279;

DT 01-FEB-2002 (first entry)

DE Peptide #2930 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast; disease; cancer.

XX Homo sapiens.

PN WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 13247; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining prediosition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;  
Best Local Similarity 99.7%; Pred. No. 4.1e-140;  
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGEAYSDNPIIRHASIANKWKTI 588

Db 1 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGEAYSDNPIIRHASIANKWKTI 60

QY 589 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 648

Db 61 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 120

QY 649 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNSLLPKDAQLSLNTLDLQL 708

Db 121 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNSLLPKDAQLSLNTLDLQL 180

QY 709 EHGDTLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 768

Db 181 EHGDTLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 240

QY 769 LGVSRERLQRTFPVSVKNGHDQGNPPDLDTTARFRVETHQKTIGGNTVKEKPPSL 828

Db 241 LGVSRERLQRTFPVSVKNGHDQGNPPDLDTTARFRVETHQKTIGGNTVKEKPPSL 300

QY 829 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 888

Db 301 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 360

QY 889 YFLDLLD 895

Db 361 YFQDLLD 367

RESULT 2

AAW56268

ID AAW56268 standard; Protein; 367 AA.

XX AC AAW56268;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28373.

XX

KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer.  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -

XX Example 4; SEQ ID NO: 28373; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is a protein encoded by one of  
XX the probes of the invention.

XX Sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;  
Best Local Similarity 99.7%; Pred. No. 4.1e-140;  
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGEAYSDNPIIRHASIANKWKTI 588

Db 1 HFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVGEAYSDNPIIRHASIANKWKTI 60

QY 589 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 648

Db 61 HLIHSGMNATTIHFNLTFQNTNDEEFKMOITVEVDTRGPKLNSTAKGYENLVSPITL 120

QY 649 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNSLLPKDAQLSLNTLDLQL 708

Db 121 LPEAEILFEDIPKEKRFKRDVNSTRAQEEVKIPLVNSLLPKDAQLSLNTLDLQL 180

QY 709 EHGDTLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 768

Db 181 EHGDTLKGYNLSKALLRSFLMNSQHAKIKKQAIITDETNDLSLVAQEKQVHKSILPNS 240

QY 769 LGVSRERLQRTFPVSVKNGHDQGNPPDLDTTARFRVETHQKTIGGNTVKEKPPSL 828

Db 241 LGVSRERLQRTFPVSVKNGHDQGNPPDLDTTARFRVETHQKTIGGNTVKEKPPSL 300

QY 829 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 888

Db 301 IVPLESQMTKEKKITGKEKENSMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 360

QY 889 YFLDLLD 895

Db 361 YFQDLLD 367



|||||  
 61 HLHSGMNAITIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAKQGYENLVSPITL 120  
 649 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 708  
 121 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 180  
 709 EHGDTITLKGYNLSKALLRSFLMNSQAHAKIKNQAIITDETNDLSLVAPOEKQVHKSIPLNS 768  
 181 EHGDTITLKGYNLSKALLRSFLMNSQAHAKIKNQAIITDETNDLSLVAPOEKQVHKSIPLNS 240  
 769 LGVSRERLQRTFFPAVSVKVNHDGQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 828  
 241 LGVSRERLQRTFFPAVSVKVNHDGQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 300  
 829 IVPLESQMTKEKKITGKEKENSMEENAHGIVTEVLLGRKLQHYTDSYLGFLPWEKKK 888  
 301 IVPLESQMTKEKKITGKEKENSMEENAHGIVTEVLLGRKLQHYTDSYLGFLPWEKKK 360  
 889 YFLDLLD 895  
 361 YFQDLLD 367

## RESULT 5

AAM04186  
 ID AAM04186 standard; Protein; 367 AA.

XX AC AAM04186;

XX DT 09-OCT-2001 (first entry)

XX DE Peptide #2868 encoded by probe for measuring breast gene expression.

XX KW Probe; human; breast disease; breast cancer; development disorder;

XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX OS Homo sapiens.

XX PN WO200157270-A2.

XX PD 09-AUG-2001.

XX PF 29-JAN-2001; 2001WO-US00661.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-476286/51.

XX DR Novel single exon nucleic acid probe used to measuring gene expression

XX PT in a human breast -

XX PS Claim 27; SEQ ID No 12926; 322pp; English.

XX CC The present invention relates to novel single exon nucleic acid probes  
 (see AAT00010-AA110067). The present sequence is a peptide encoded by one  
 such probe. The probes are useful for measuring human gene expression in  
 a human breast sample, where the probe hybridises at high stringency to a  
 nucleic acid expressed in the human breast. The probes are useful for  
 predicting, diagnosing, grading, staging, monitoring and prognosing  
 diseases of the human breast, particularly those diseases with polygenic  
 aetiology. The diseases include: breast cancer, disorders of polyploid  
 inflammatory diseases of the breast, fibrocystic changes, proliferative

CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ sequence 367 AA;

Query Match 29.9%; Score 1893; DB 22; Length 367;  
 Best Local Similarity 99.7%; Pred. No. 4.le-140;  
 Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 529 HFHELYKVVLLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIAKWKTI 588  
 DB 1 HFHELYKVVLLPNQTHYIIPKGECLPYFSFAEVAKRGVEGAYSDNPIIRHASIAKWKTI 60  
 QY 589 HLIMHSGMNAITIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAKQGYENLVSPITL 648  
 DB 61 HLIMHSGMNAITIHFNLTQNTNDEEFKQITVEVDTRGPKLNSTAKQGYENLVSPITL 120  
 QY 649 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 708  
 DB 121 LPEAEILFEDIPKEKRPKPKRHDVNSTRAQEVKIPLVNISLLPKDAQLSLNTLDLQ 180  
 QY 709 EHGDTITLKGYNLSKALLRSFLMNSQAHAKIKNQAIITDETNDLSLVAPOEKQVHKSIPLNS 768  
 DB 181 EHGDTITLKGYNLSKALLRSFLMNSQAHAKIKNQAIITDETNDLSLVAPOEKQVHKSIPLNS 240  
 QY 769 LGVSRERLQRTFFPAVSVKVNHDGQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 828  
 DB 241 LGVSRERLQRTFFPAVSVKVNHDGQGNPDLDTTARFRVETHQKTIGGNVTKEKPPSL 300  
 QY 829 IVPLESQMTKEKKITGKEKENSMEENAHGIVTEVLLGRKLQHYTDSYLGFLPWEKKK 888  
 DB 301 IVPLESQMTKEKKITGKEKENSMEENAHGIVTEVLLGRKLQHYTDSYLGFLPWEKKK 360  
 QY 889 YFLDLLD 895  
 DB 361 YFQDLLD 367

## RESULT 6

ABG38224

ID ABG38224 standard; Peptide; 367 AA.

XX AC ABG38224;

XX DT 19-AUG-2002 (first entry)

XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 27889.

XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;

XX KW chronic obstructive pulmonary disease; interstitial lung disease;

XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;

XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX KW primary ciliary dyskinesia; pulmonary hypertension;

XX KW hyaline membrane disease.

XX XX Homo sapiens.

XX PN WO200186003-A2.

XX XX 15-NOV-2001.

XX PD 30-JAN-2001; 2001WO-US00665.

XX PF 04-FEB-2000; 2000US-180312P.

XX PR 26-MAY-2000; 2000US-207456P.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-234687P.

PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2002-114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
XX Claim 27; SEQ ID NO 27889; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC and hyaline membrane disease. The present sequence is a peptide/protein  
CC encoded by a single exon probe of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 367 AA;  
XX  
XX Query Match 29.9%; Score 1893; DB 23; Length 367;  
XX Best Local Similarity 99.7%; Pred. No. 4.1e-140;  
XX Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
QY 529 HFHELYKVILLNTHYIIPKGECLPYFSFAEVARGVGAGVSDNPPIIRHASIAKWKTI 588  
DB 1 HFHELYKVILLNTHYIIPKGECLPYFSFAEVARGVGAGVSDNPPIIRHASIAKWKTI 60  
QY 589 HLTMHSGMNATTIFHNLTFQNTNDEEFKMQITVEVDTRGPKLNSTAGKYNLSPITL 648  
DB 61 HLTMHSGMNATTIFHNLTFQNTNDEEFKMQITVEVDTRGPKLNSTAGKYNLSPITL 120  
QY 649 LPFAEILFEDIPKRFKFKHVDNSTRRAQEEVKIPLVNISLLPKDAQLSNTLDLQL 708  
DB 121 LPFAEILFEDIPKRFKFKHVDNSTRRAQEEVKIPLVNISLLPKDAQLSNTLDLQL 180

QY 709 EHGDTLKGYNLSKALLRSFLMNSOHAHAKIKNOAILTDETNDLSLVAPOEKQVHKSILPNS 768  
DB 181 EHGDTLKGYNLSKALLRSFLMNSOHAHAKIKNOAILTDETNDLSLVAPOEKQVHKSILPNS 240  
QY 769 LGVSRERQLRTPPAVSVKVNGHDQGNPPLDLETTARFRVETHQTKTIGGNVTKPPSL 828  
DB 241 LGVSRERQLRTPPAVSVKVNGHDQGNPPLDLETTARFRVETHQTKTIGGNVTKPPSL 300  
QY 829 IVPLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 888  
DB 301 IVPLESQMTKEKKITGKEKENSMEENAENHIGVTEVLLGRKLQHYTDSYLGFLPWEKKK 360  
QY 889 YFLDLLD 895  
DB 361 YFQDLLD 367  
RESULT 7  
ABB59094  
ID ABB59094 standard; Protein; 652 AA.  
XX  
AC ABB59094;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 4074.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL03197.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 4074; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 652 AA;  
XX  
XX Query Match 13.5%; Score 854.5; DB 22; Length 652;  
XX Best Local Similarity 24.4%; Pred. No. 6e-58;  
XX Matches 220; Conservative 107; Mismatches 201; Indels 373; Gaps 15;  
QY 305 DEDISARFEDNEELRYSLRHSIERHAPVWRNIFVINGQIPSWNLNDNPRVTIVTHQDVP 364





AC ABB38187;  
XX  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #5693 encoded by human foetal liver single exon probe.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 30822; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 68 AA;  
SQ  
Query Match 5.6%; Score 352; DB 22; Length 68;  
Best Local Similarity 100.0%; Pred. No. 4.8e-20;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 113 EKQECLLTHCIVKPMVLDPALPANITLKDLPSPFHSASDIFNVAKPKNPSTNVSV 172  
Db 1 EKQECLLTHCIVKPMVLDPALPANITLKDLPSPFHSASDIFNVAKPKNPSTNVSV 60  
QY 173 VVFDSTKD 180  
Db 61 VVFDSTKD 68  
RESULT 10  
AAM58821  
ID AAM58821 standard; Protein; 68 AA.  
XX  
XX AAM58821;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 30926.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.

XX  
OS Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
XX Example 4; SEQ ID NO: 30926; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
XX Sequence 68 AA;  
SQ  
Query Match 5.6%; Score 352; DB 22; Length 68;  
Best Local Similarity 100.0%; Pred. No. 4.8e-20;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 113 EKQECLLTHCIVKPMVLDPALPANITLKDLPSPFHSASDIFNVAKPKNPSTNVSV 172  
Db 1 EKQECLLTHCIVKPMVLDPALPANITLKDLPSPFHSASDIFNVAKPKNPSTNVSV 60  
QY 173 VVFDSTKD 180  
Db 61 VVFDSTKD 68  
RESULT 11  
AAM71337  
ID AAM71337 standard; Protein; 68 AA.  
XX  
XX AAM71337;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 31643.  
XX  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
XX Homo sapiens.  
XX  
XX WO200157276-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00668.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR

PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488900/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 31643; 658pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 XX  
 SQ Sequence 68 AA;  
 Query Match 5.6%; Score 352; DB 22; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-20;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 113 EKQLECLLTHCICKVPMVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPKNPSTNVSV 172  
 Db 1 EKQLECLLTHCICKVPMVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPKNPSTNVSV 60  
 Qy 173 VVFDSTKD 180  
 Db 61 VVFDSTKD 68  
 RESULT 12  
 AAM31616  
 ID AAM31616 standard; Protein; 68 AA.  
 AC AAM31616;  
 DT 17-OCT-2001 (first entry)  
 XX Peptide #5653 encoded by probe for measuring placental gene expression.  
 DE Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 KW Homo sapiens.  
 OS WO200157272-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US00663.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -

DR WPI; 2001-48897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 PS Claim 27; SEQ ID No 31885; 654pp; English.  
 XX The present invention relates to single exon nucleic acid probes (SENP;  
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX  
 SQ Sequence 68 AA;  
 Query Match 5.6%; Score 352; DB 22; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-20;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 113 EKQLECLLTHCICKVPMVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPKNPSTNVSV 172  
 Db 1 EKQLECLLTHCICKVPMVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPKNPSTNVSV 60  
 Qy 173 VVFDSTKD 180  
 Db 61 VVFDSTKD 68  
 RESULT 13  
 ABG41140  
 ID ABG41140 standard; Peptide; 68 AA.  
 AC ABG41140;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 30805.  
 DE Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX Homo sapiens.  
 OS WO200186003-A2.  
 PN 15-NOV-2001.  
 PD 30-JAN-2001; 2001WO-US00665.  
 PF 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -

XX PS Claim 27; SEQ ID NO 30805; 634pp; English.

XX CC The invention relates to a spatially-addressable set of single exon

XX CC nucleic acid probes for measuring gene expression in a sample derived

XX CC from human lung comprising single exon nucleic acid probes having one of

XX CC 12614 nucleic acid sequences mentioned in the specification, or their

XX CC complements or the 12387 open reading frames derived from the 12614

XX CC probes. Also included are a microarray comprising the novel set of

XX CC probes; the novel set of probes which hybridise at high stringency to a

XX CC nucleic acid expressed in the human lung; measuring gene expression in a

XX CC sample derived from human lung, comprising (a) contacting the array with

XX CC a collection of detectably labeled nucleic acids derived from human lung

XX CC mRNA, and (b) measuring the label detectably bound to each probe of

XX CC the array; identifying exons in a eukaryotic genome, comprising

XX CC (a) algorithmically predicting at least one exon from genomic sequences

XX CC of the eukaryote; and (b) detecting specific hybridisation of detectably

XX CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

XX CC having a fragment identical to the predicted exon, the probe is included

XX CC in the above mentioned microarray; assigning exons to a single gene,

XX CC comprising (a) identifying exons from genomic sequence by the method

XX CC above and (b) measuring the expression of each of the exons in several

XX CC tissues and/or cell types using hybridisation to a single exon

XX CC microarrays having a probe with the exon, where a common pattern of

XX CC expression of the exons in the tissues and/or cell types indicates that

XX CC the exons should be assigned to a single gene; a peptide comprising one

XX CC of 12011 sequences, mentioned in the specification, or encoded by the

XX CC probes/open reading frames (ORF). The probes are used for gene

XX CC expression analysis, and for identifying exons in a gene, particularly

XX CC using human lung derived mRNA and for the study of lung diseases

XX CC such as asthma, lung cancer, chronic obstructive pulmonary disease

XX CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

XX CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

XX CC Niemann-Pick disease, Hermansky-Budlak syndrome, sarcoidosis, pulmonary

XX CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,

XX CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic

XX CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

XX CC and hyaline membrane disease. The present sequence is a peptide/protein

XX CC encoded by a single exon probe of the invention.

XX CC Note: The sequence data for this patent did not form part

XX CC of the printed specification, but was obtained in electronic

XX CC format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 68 AA;

Query Match 5.6%; Score 352; DB 23; Length 68;

Best Local Similarity 100.0%; Pred. No. 4.8e-20;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 EKQECCLTHCIKVPMLVDPALPANITLKDLPSPFHSASDIFNVAKPNPSTNSV 172

Db 1 EKQECCLTHCIKVPMLVDPALPANITLKDLPSPFHSASDIFNVAKPNPSTNSV 60

QY 173 VVFDSTKD 180

Db 61 VVFDSTKD 68

RESULT 14

AAW79296

ID AAW79296 standard; Protein; 545 AA.

XX AC AAW79296;

XX DT 15-FEB-1999 (first entry)

XX DE Neisseria UDP-N-acetyl-mannosamine (alpha-6) polymerase.

XX KW Serogroup marker; molecular typing; capsular instability; vaccine;

XX KW meningitis; UDP-N-acetyl-mannosamine (alpha-6) polymerase.

XX OS Neisseria meningitidis serogroup A strain F8229.

XX PN WO9845312-A1.

XX PD 15-OCT-1998.

XX PF 09-APR-1998; 98WO-US06946.

XX PR 23-SEP-1997; 97US-0936107.

XX PR 09-APR-1997; 97US-0069885.

XX PA (UYEM-) UNIV EMORY.

XX PI Stephens DS, Swartley JS;

XX DR WPI; 1998-583189/49.

XX DR N-PSDB; AAW69110.

XX PT New DNA sequences from *Neisseria meningitidis* involved in serotype A

XX PT polysaccharide synthesis - and recombinant sequences encoding

XX PT capsule-switching mutation(s) in *N. meningitidis*, for producing

XX PT polysaccharides useful in vaccines

XX PS Disclosure; Page 55-59; 111pp; English.

XX CC This polypeptide is encoded by ORF2, one of 4 open reading frames

XX CC identified in the region between the *ctrA* and *galE* genes of the

XX CC encapsulated serogroup A *Neisseria meningitidis* strain F8229.

XX CC The ORF2 protein is predicted to be a UDP-N-acetyl-mannosamine

XX CC (alpha1-6) polymerase involved in the biosynthesis of the serogroup

XX CC A capsule. The ORF1 and ORF3-ORF4 gene products (see AAW79295 and

XX CC AAW79297-98) are also thought to be involved in capsule biosynthesis or

XX CC modification. The invention provides a model in which meningococcal

XX CC capsular serogroups are determined by specific genetic biosynthesis

XX CC cassettes that insert between the *ctrA* operon and *galE*. For

XX CC serogroup A, the cassettes determining specificity of serogroups

XX CC can recombine to switch the type of capsule and serogroup expressed.

XX CC Such information is critical to the design of improved group A and

XX CC other meningococcal vaccines and to the understanding of the

XX CC molecular basis of serogroup A pathogenesis. The invention

XX CC provides recombinant meningococcal strains, recombinant DNA

XX CC constructs and immunological preparations useful as diagnostic

XX CC probes for detection and diagnosis of meningococcal diseases,

XX CC screening for specific serogroups and broad-based immunisations

XX CC with multivalent capsular polysaccharide conjugate vaccines.

XX SQ Sequence 545 AA;

Query Match 4.2%; Score 264; DB 19; Length 545;

Best Local Similarity 27.8%; Pred. No. 1.3e-11;

Matches 90; Conservative 52; Mismatches 130; Indels 52; Gaps 13;

QY 135 LPANITLKDLPSPFHSASDIFNVAKPNPSTNSVVFSDTKDVEDAHSGLK-G-NS 193

Db 97 LPSNLTUK--PALCILESHKEDFLN-----KFLLTISSENKLYQFNQIRNPKS 145

QY 194 RQTVWRCY-----LTTDKVEPGLV-----LMQDLAPLSGFPPTFKETNOLKT 235

Db 146 VNEIWTDLFSIAHVDMKLTDRTLSSISQFWFRLEPCKEDKDFI-----LFTANRYSR 200

QY 236 KLPENLSKVKLLQLYSEASVALLKLNPKDFQELNKQTKNNITDGKELTISPAYLLWD 295

Db 201 KLWKH---SIKNNOLFKEG-----IRNYSEISLSPYEEDHNFIDILFTWVNSEDKNQW 251

QY 296 LSAISOSKQD---EDISARFEDNEELRYSLRSTERHAPVNRNIFVTNQIIPSWLNDN 352

Db 252 -ELYKKYKDPFNSDATSTSRFLSRDELKFAIRSWEMSGSFIRKIFIVSNCAPAWLDLN 310

QY 353 PRVTIVTHQDVFRNLSHLPFSPAIESHVRIEGLSQKFTYLLMDVDFGKQVWPDFFYS 412

Db 311 PKIQWVYHEEIMPQ-SALPTFSSHAIEITSLHHIPGISNYFIYSNDDFLTKPLNKDNFF 369

QY 413 HSKGQKVYL-TWVPVN--CAEGCP 433

Db 370 SNGIAKLRLKLEAWGNVNGECTEGEP 393

RESULT 15

AAE18207

ID AAE18207 standard; Protein: 2469 AA.

XX AC AAE18207;

XX DT 07-MAY-2002 (first entry)

XX DE Human MOLLA protein.

XX Secreted molecule; MOLLA protein; MOLX; cardiomyopathy; atherosclerosis; diabetes; chromosomal disorder; albinism; anaemia; psoriasis; scarring; liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection; cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma; immune disease; ischaemia; immunodeficiency; rheumatoid arthritis; ulcer; HIV; human immuno deficiency virus; hepatitis; haemostatic disease; pain; haematopoietic; thrombolytic; thrombocytopaenia; Alzheimer's disease; Parkinson's disease; Huntington's disease; muscular disease; stress; ocular disease; growth disorder; depression; epilepsy; contraceptive; vulnery; osteopathic; haemostatic; tranquiliser; antidepressant; analgesic; vasotropic; hypotensive; gene therapy.

XX OS Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..25

XX Protein /label= Signal\_peptide

XX /label= Mature\_MOLLA\_protein

XX W0200206339-A2.

XX 24-JAN-2002.

XX 03-JUL-2001; 2001WO-US21249.

XX 03-JUL-2000; 2000US-215854P.

XX 03-JUL-2000; 2000US-215856P.

XX 03-JUL-2000; 2000US-215902P.

XX 07-JUL-2000; 2000US-216585P.

XX 07-JUL-2000; 2000US-216586P.

XX 07-JUL-2000; 2000US-216722P.

XX 17-JUL-2000; 2000US-218622P.

XX 17-JUL-2000; 2000US-218992P.

XX 27-JUL-2000; 2000US-221285P.

XX 14-FEB-2001; 2001US-268734P.

(CURA-) CURAGEN CORP.

XX Spaderna SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;

XX Patturajan M, Taupier RJ, Rastelli L, Grosse WM, Szekeres ES;

XX Alsbrook J, Lepley DM, Shen L, Burgess CE, Shimkets RA;

XX Padigar M;

XX WPI; 2002-155038/20.

XX N-PSDB; AAD28941.

XX Nucleic acids encoding secreted polypeptides, designated MOLX

XX polypeptides, useful for treating a MOLX-associated disorder, e.g.

XX cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -

XX Claim 1; Page 11; 223pp; English.

XX The patent discloses nucleic acid sequences encoding novel secreted

XX molecule (MOL) polypeptides, designated MOLX polypeptides (i.e. a MOL

XX protein where X is an integer from 1 to 8). Sequences of the invention

XX are useful for treating or preventing a MOLX-associated disorder in

XX humans. They are useful for treating or preventing cardiomyopathy,

XX atherosclerosis and disorders related to cell signal processing and

XX metabolic pathway modulation. The MOLX antibodies are useful for

CC treating or preventing diabetes and disorders related to cell signal

CC processing and metabolic pathway modulation. MOLX sequences are useful

CC for the treatment or diagnosis of other MOLX-associated disorders, e.g.

CC chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,

CC scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral

CC thrombosis, ischaemia, hypertension, systemic lupus erythematosus, immune

CC diseases, immunodeficiency, HIV (human immuno deficiency virus), viral,

CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,

CC haematopoietic, thrombolytic, haemostatic diseases, thrombocytopaenia,

CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,

CC ocular disease, muscular diseases, growth disorders, loss of libido,

CC stress, depression, pain and epilepsy. They are useful for preventing

CC chemotherapy side effects and as contraceptives. Sequences of the

CC invention are also useful for gene therapy. The present sequence

XX is human Notch-like protein, MOLLA.

SQ Sequence 2469 AA;

Query Match 2.7%; Score 171.5; DB 23; Length 2469;

Best Local Similarity 19.9%; Pred. No. 0.0033;

Matches 111; Conservative 63; Mismatches 176; Indels 207; Gaps 26;

QY 424 PVPNCAEGCPGSMWIKDYCDKACNNSACDWDGDCSGNNGSRYTAGGGTSGIGVGPW 483

DB 1419 PPATCLSQYCADKARDGVCDEACNSHACQWDGDC-----SUTMENPW 1461

QY 484 -QFGGINSVSYCNQGC-----ANS-----WLADKF-----CDQACN 514

DB 1462 ANCSPLPCWDYINNQCDELNTVECLDFNFCQGNSTCKYKACADHFNCDQGCN 1521

QY 515 VLSCGFADGDCQDFHELYK-----VILLPNQTHIIPKGECLPYFSFAEV-----561

DB 1522 SEECGWDGLCAADQCPENLAEGTLVIVLMP-----PEQLLOARSFLRALTLLHN 1574

QY 562 --AKRGEVAGSDNPPIIRHASTANKKWTIHLIMHSGMNATTIHFNLTFQNTDEEFK---616

DB 1575 LRIKRDQSGELMVYPYGEKSAAMKKQ-----RMTRRSLPGQEQEVA 1617

QY 617 -MQITVEVDTREGPKLNSTAQKGYENLYSPITLLPEAEILFEDIPKRFKFRHDVNS 675

DB 1618 GSKVFEIDNRQ---CVQDSHCFKNTDAAALLAS-----HAIQG 1655

QY 676 TRRAQEEVKIPLVNI---SLLPKDAQLSLNTLDLQL-----EHGDIIT 714

DB 1656 T-----LSYPLVSVVSESLTPERTQL-LYLLAVAVVILFIILLGYIMAKRKRKHSGLW 1708

QY 715 L-KGYNLSKALLRSFLMNSQAK-----IKNOAITDETNSLVAPOEKQVHKS 763

DB 1709 LPEGFTLRDA-----SNHKKREPVGODAVGLKNLSQVSEAN--LI-----1748

QY 764 ILPNSLGVSERLQRLTFPAVSVKYNHGDG-----QNPPLDLETTARFRVETHQKTIGN 819

DB 1749 ---GTGTSEHWVDEGFPQK-KYKADEALLSEDDPIDR-----RPWTOQHLEAA 1795

QY 820 VTKKPPSLVPLPESQMTKE-----KKITGKEKENSMEENAEH---H 859

DB 1796 DIRRTPSLALTPQAEQEVLDVNVVRGPDGCTPLMLASLRGSSDLSDDEDAEDSSAN 1855

QY 860 IGVTEVLLGRKLOHYTD 876

DB 1856 IITDLIVQGASLQAQTD 1872

Search completed: April 13, 2003, 03:20:30

Job time : 75.7772 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 03:00:01 ; Search time 63.7872 Seconds  
(without alignments)  
3873.042 Million cell updates/sec

Title: US-10-023-888-2  
Perfect score: 6340  
Sequence: 1 METDTLLWLLVWPGSTG.....NRFLMHLEQWRAYRDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_podent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4248	67.0	847	4 Q96N13	Q96n13 homo sapien
2	3127	49.3	663	4 Q9UL12	Q9ul12 homo sapien
3	1281.5	20.2	384	11 Q6L1340	Q6l1340 mus musculus
4	1046	16.5	248	4 Q9NPW9	Q9npw9 homo sapien
5	854.5	13.5	652	5 Q9V553	Q9v553 drosophila
6	854.5	13.5	666	5 Q8SX14	Q8sx14 drosophila
7	460	7.3	132	4 Q9NV34	Q9nv34 homo sapien
8	327	5.2	602	16 Q9L112	Q9l112 streptomyce
9	324.5	5.1	942	16 Q9L114	Q9l114 streptomyce
10	321.5	5.1	586	16 Q69853	Q69853 streptomyce
11	313	4.9	541	16 Q69851	Q69851 streptomyce
12	286	4.5	238	2 Q9EVX1	Q9evx1 streptococc
13	265	4.2	545	16 Q9JW8	Q9jw8 neisseria m
14	264	4.2	545	2 Q68215	Q68215 neisseria m
15	259.5	4.1	442	5 Q8SSQ9	Q8ssq9 dictyosteli
16	255	4.0	532	16 Q06628	Q06628 mycobacteri

17	253.5	4.0	542	16 Q50025	O50025 mycobacteri
18	229.5	3.6	366	2 Q9RGR0	Q9rgr0 neisseria m
19	227.5	3.6	373	2 Q51151	Q51151 neisseria m
20	190	3.0	1476	13 Q90285	Q90285 carassius a
21	184	2.9	2752	5 Q9BJY0	Q9bjy0 plasmodium
22	183	2.9	2867	5 Q9N2M3	Q9n2m3 plasmodium
23	182	2.9	638	13 Q42372	Q42372 brachydanio
24	180.5	2.8	2352	5 Q61240	Q61240 halocynthia
25	176	2.8	1422	5 Q00914	Q00914 plasmodium
26	176	2.8	1999	11 Q63731	Q63731 rattus norv
27	171	2.7	2524	5 Q9GPA5	Q9gpa5 brachioosto
28	169.5	2.7	762	13 Q42373	Q42373 brachydanio
29	167.5	2.6	523	5 Q9NE33	Q9ne33 leishmania
30	167.5	2.6	3144	5 Q9GTK4	Q9gtk4 plasmodium
31	167	2.6	1387	5 Q9GZ76	Q9gz76 plasmodium
32	166	2.6	3130	5 Q9BK46	Q9bk46 plasmodium
33	164	2.6	3254	5 Q9BK45	Q9bk45 plasmodium
34	163.5	2.6	763	16 Q8XJY4	Q8xjy4 clostridium
35	163.5	2.6	2771	5 Q26216	Q26216 plasmodium
36	162.5	2.6	1979	5 Q96133	Q96133 plasmodium
37	162	2.6	2434	8 Q9MTH5	Q9mth5 oenothera h
38	161.5	2.5	898	11 Q9Z220	Q9z220 rattus norv
39	161.5	2.5	1694	5 Q9T7T5	Q9t7t5 plasmodium
40	160.5	2.5	1694	5 Q9NHX1	Q9nhx1 plasmodium
41	160.5	2.5	1720	5 Q25922	Q25922 plasmodium
42	160.5	2.5	3259	4 Q14789	Q14789 homo sapien
43	159.5	2.5	1928	5 Q9U0H2	Q9u0h2 plasmodium
44	158.5	2.5	2104	16 Q8XM24	Q8xm24 clostridium
45	158	2.5	1194	13 Q9W737	Q9w737 gallus gall

## ALIGNMENTS

## RESULT 1

Q96N13 ID Q96N13 PRELIMINARY; PRT; 847 AA.  
AC Q96N13;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE CUNA FLJ31575 fis, clone NT2R12001846, moderately similar to basic domain/leucine zipper transcription factor (Fragment).  
DE Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H., Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamanoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
RA "NEDO human cDNA sequencing project."  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
EMBL; AK056137; BAB71102.1;  
DR InterPro: IPR000800; Notch.  
DR Pfam: PF00066; notch; 1.  
FT NON\_TER 847 847  
SQ SEQUENCE 847 AA; 95165 MW; 0A115015824733C5 CRC64;

Query Match 67.0%; Score 4248; DB 4; Length 847;  
Best Local Similarity 99.9%; Pred. No. 11e-245;  
Matches 802; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 SRDQYHVLFDSDYRDNTAGKSFQNRCLCLPMPIDVVYTWNGTDLLELKLQVREQMEERQ 94  
|||||  
Db 45 SRDQYHVLFDSDYRDNTAGKSFQNRCLCLPMPIDVVYTWNGTDLLELKLQVREQMEERQ 104  
|||||

```
QY 95 KAMEILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITIKDLPSPYSEFHS 154
DB 105 KAMEILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDPALPANITIKDLPSPYSEFHS 164
QY 155 SDIPNAKPNPSTNVSVVYFDSSTKQVEDAHSGILLKGNRSQTVWGYLTDTKEVPGVL 214
DB 165 SDIPNAKPNPSTNVSVVYFDSSTKQVEDAHSGILLKGNRSQTVWGYLTDTKEVPGVL 224
QY 215 QDLAFLSGFPPTFKETNOLATKIPENLSKVLLQLYSEASVALLKLNPKDFQELNKOT 274
DB 225 QDLAFLSGFPPTFKETNOLATKIPENLSKVLLQLYSEASVALLKLNPKDFQELNKOT 284
QY 275 KKNMTIDGKELTISPAYLLWDLISAISQSKODEDISASRFEDNEELRYSLSIERHAPWY 334
DB 285 KKNMTIDGKELTISPAYLLWDLISAISQSKODEDISASRFEDNEELRYSLSIERHAPWY 344
QY 335 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVPRNLNLSHLPFTSSPAIESHVHRIEGLSKFTY 394
DB 345 NIFIVTNGQIPSWNLNDNPRVTIVTHQDVPRNLNLSHLPFTSSPAIESHVHRIEGLSKFTY 404
QY 395 LNDVMPFGKDWPDYFHSKGGKYLTPVPCNAECGPGSWIKDGYCDKACNNSACDWD 454
DB 405 LNDVMPFGKDWPDYFHSKGGKYLTPVPCNAECGPGSWIKDGYCDKACNNSACDWD 464
QY 455 GDCSGNSGGSRYTAGGGTGSIGVGPWQFGGINSVSYCNQGCANSLWADKFCDAQCN 514
DB 465 GDCSGNSGGSRYTAGGGTGSIGVGPWQFGGINSVSYCNQGCANSLWADKFCDAQCN 524
QY 515 VLSGCFDAGCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNP 574
DB 525 VLSGCFDAGCGQDHFHLYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNP 584
QY 575 IIRASIANKWKTHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRREGPKLNST 634
DB 585 IIRASIANKWKTHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRREGPKLNST 644
QY 635 AOKGYENLVSPITLPPAEILFEDIPKEKRPKFKRHDVNSTRAEQEVKIPILVNISL 694
DB 645 AOKGYENLVSPITLPPAEILFEDIPKEKRPKFKRHDVNSTRAEQEVKIPILVNISL 704
QY 695 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVA 754
DB 705 KDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVA 764
QY 755 PQEKQVHKSILPNSLGVSRQLRTPFAVSVKNGHDQGNPDLLETTARFRVETHQK 814
DB 765 PQEKQVHKSILPNSLGVSRQLRTPFAVSVKNGHDQGNPDLLETTARFRVETHQK 824
QY 815 TIGGNVTKKPPSLIVPLESQMT 837
DB 825 TIGGNVTKKPPSLIVPLESQMT 847

RESULT 2
Q9ULL2 PRELIMINARY; PRT; 663 AA.
AC Q9ULL2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE K1A1208 protein (Fragment).
GN K1A1208.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
```

```
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL: AB033034; BAA86522.1; -.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 663 AA; 76150 MW; 76B4B46242C2CA8 CRC64;

Query Match 49.3%; Score 3127; DB 4; Length 663;
Best Local Similarity 98.5%; Pred. No. 8.7e-179; Indels 0; Gaps 0;
Matches 603; Conservative 2; Mismatches 7;

QY 528 DHFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPPIIRHASIANKWK 587
DB 22 DHFHELYKVILLPNQTHYIIPKGECLPYFSAEVAKRGVEGAYSDNPPIIRHASIANKWK 81
QY 588 IHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRREGPKLNSTRAEQEVKIPILVN 647
DB 82 IHLIMHSGMNTTIFNLTFQNTNDEEFKQITVEVDTRREGPKLNSTRAEQEVKIPILVN 141
QY 648 LPEAEILFEDIPKEKRPKFKRHDVNSTRAEQEVKIPILVNISLPPKDAQLSLNTLDLQ 707
DB 142 LPEAEILFEDIPKEKRPKFKRHDVNSTRAEQEVKIPILVNISLPPKDAQLSLNTLDLQ 201
QY 708 LEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVAPOEKQVHKSILPN 767
DB 202 LEHGDITLKGYNLSKALLRSFLMNSQAKIKNOAIITDETNDLSVAPOEKQVHKSILPN 261
QY 768 SLGVSRQLRTPFAVSVKNGHDQGNPDLLETTARFRVETHQKIGGNVTKKPPS 827
DB 262 SLGVSRQLRTPFAVSVKNGHDQGNPDLLETTARFRVETHQKIGGNVTKKPPS 321
QY 828 LTVPLESQMTKEKTKITGKEKENSMEENAHIGVTEVLLGRKLOHYTDSYLGFLPWEKK 887
DB 322 LTVPLESQMTKEKTKITGKEKENSMEENAHIGVTEVLLGRKLOHYTDSYLGFLPWEKK 381
QY 888 KYFDLLDEESLKQLAYFTDTSKNRKYKRDTPADSLRYVKNILNSFGFTSRKVPAMH 947
DB 382 KYFDLLDEESLKQLAYFTDTSKNRKYKRDTPADSLRYVKNILNSFGFTSRKVPAMH 441
QY 948 PHMDIRYVQELQDMPEEFDKTSFKVHSEDMQFAFSFYFLMSAVQPNISQVDFEV 1007
DB 442 PHMDIRYVQELQDMPEEFDKTSFKVHSEDMQFAFSFYFLMSAVQPNISQVDFEV 501
QY 1008 DTDSQSVLSDEIRTLATRIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIPPTQES 1067
DB 502 DTDSQSVLSDEIRTLATRIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIPPTQES 561
QY 1068 YTDPNLPPVTKSLVNTCKPVTDKIHKAYKDKNRYRFEIMGEEETAFKMRITNVSHVQQL 1127
DB 562 YTDPNLPPVTKSLVNTCKPVTDKIHKAYKDKNRYRFEIMGEEETAFKMRITNVSHVQQL 621

QY 1128 DDIRKNPKFVC 1139
DB 622 DDIRKNPKFVC 633

RESULT 3
Q61340 PRELIMINARY; PRT; 384 AA.
ID Q61340
AC Q61340;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Basic domain/leucine zipper transcription factor (Fragment).
GN MAPB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scurionathia; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=95094266; PubMed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RL zipper transcription factor.";
RL Cell 79:1025-1034(1994);
DR EMBL: L36434; AAA65688.1; -.
DR MGD: MGI:104555; Matf.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 384 AA; 43754 MW; 51F473C8807A7E55 CRC64;

Query Match
Best Local Similarity 67.8%; Score 1281.5; DB 11; Length 384;
Matches 271; Conservative 31; Mismatches 77; Indels 21; Gaps 6;

QY 634 TAQKGYENLYSPITLLPENAELFEDIPKEKRPKRDHVNSTRRAQEVKIPLVNISLL 693
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3 TTQKAYESLVSPVTPLPQADVPFEDVPKRPKIRRHVDVATGFRQBEVKIPRVNISLL 62

QY 694 PKDAQSLNTLDLQLEHGDTITLKGYNLSKALLRSLFNLMSQAKTKNOAILTDEINDSLV 753
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
63 PKEAQVRLNLDLQLEHGDTITLKGYNLSKALLRSLFNLMSQAKTKNOAILTDEINDSLV 121

QY 754 APOEKQVHKSILPLSLGVSRQLRTPFPAVSKVNGHDQGNPPDLDTETARFRVETHTQ 813
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
122 VPQENPSHRR--PHGFAGEHRSERWTAETAETVTVKGRDHALNPPVPLETNARL-----AQ 174

QY 814 KITGNGVTEKPPSLIVPLESOMTEKKITGKEKENSMEENAHIGVTEVLGRKLQH 873
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
175 PTIGTVSKENISPLIVPPESHLP-----KEESDRAEGNA---VPVKELVPGRRCSR 224

QY 874 YTDVSLGFLPWEKKYFELDLDEESLKTOLAYFTDTSKNRARKDFTADSLRVYKNILN 933
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
225 ITQA---FCPGKKKKYFQDLDAESLKTOLAYFTDTSKHTGRLQKDTFADSLRVYKNILN 281

QY 934 SKFGFTSRKVPAPHPMDIRVMOQLDMQFPEEFKTSFHKVRSEDMQFAFSFYFLMS 993
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
282 SKFGFTSRKVPAPHPMDIRVMOQLDMQFPEEFKTSFHKVRSEDMQFAFSFYFLMS 341

QY 994 AVQPLNISQVFDEVTDSGVLSDEIRTLATRIHELPLS 1033
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
342 AVQPLNISQVFDEVTDSGVLSDEIRTLATRIHDLPLT 381

RESULT 4
Q9NPW9 PRELIMINARY; PRT; 248 AA.
AC Q9NPW9
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypoetical 29.3 kDa protein (Fragment).
GN DKFZP762B26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Bloeker H., Boeher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL359588; CAB94874.1; -.
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Hypoetical protein.
FT NON_TER 1
SQ SEQUENCE 248 AA; 29336 MW; DAC4C4A6C4505522 CRC64;

Query Match
Best Local Similarity 16.5%; Score 1046; DB 4; Length 248;
Matches 100.0%; Pred. No. 5.8e-55;

```

```

Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 10;

QY 1003 VFDEVDTDQSGVLSDRIRIATRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNNIP 1062
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
5 VFDEVDTDQSGVLSDRIRIATRIHELPLSLQDLTGLEHMLNCSKMLPADITQLNNIP 64

QY 1063 PTOESYDNPILPPVTKSLVTNCKPVTOKIKAYKDKNKYRFEIMGEEIAFKMIRTNVSH 1122
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
65 PTOESYDNPILPPVTKSLVTNCKPVTOKIKAYKDKNKYRFEIMGEEIAFKMIRTNVSH 124

QY 1123 VVGQLDDIRKNPRKFCVCLNDNIDHNHDKDAQTVKAVLRDFYESMFPPIPSQFELPREYRNR 1182
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
125 VVGQLDDIRKNPRKFCVCLNDNIDHNHDKDAQTVKAVLRDFYESMFPPIPSQFELPREYRNR 184

QY 1183 LHMHELOEWRAYRDKLK 1199
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
185 LHMHELOEWRAYRDKLK 201

RESULT 5
Q9V553 PRELIMINARY; PRT; 652 AA.
AC Q9V553
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG8027 protein.
GN CG8027.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anandates P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe C.R., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brokstein P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhao G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL: AE003834; AAF58967.1; -.

DR FlyBase; FBgn0033392; CG8027.

DR InterPro; IPR000800; Notch.

DR Pfam; PF00066; notch; 1.

DR SMART; SM00004; NL; 1.

SQ SEQUENCE 652 AA; 75752 MW; 2728764810039458 CRC64;

Query Match 13.5%; Score 854.5; DB 5; Length 652;

Best Local Similarity 24.4%; Pred. No. 6.7e-43;

Matches 220; Conservative 107; Mismatches 201; Indels 373; Gaps 15;

QY 305 DEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNDNPRVTIVTHQDVF 364

Db 72 DDKYDPSRFDKNEELRYSLRSLEKHAWIIRVIVTNGQIPSWLDSYERVTVPVHEVLA 131

QY 365 RNLSHLPTFSSPAIESHVRIEGLSQFIYLNDDVFMFGKVDWPDYSHSKGQKVIYLTWP 424

Db 132 PDPQLPTFSSSAIETFLHRIPKSLRFLYLNDDIFLGAPLYPEDLYTEAGVRVYQAWM 191

QY 425 VPCNAEGCPGSKWIDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGIGVQWPQ 484

Db 192 VPGCALDCPWTYIGDGACDRHCNIDACQFDGDCSETGPAS----- 232

QY 485 FGGGINSVSYCNOGCANSWLADKFCQACNVLSGDFDAGDGHFELYKVILLPNQTH 544

Db 233 -----DAH 235

QY 545 YIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHLMHSGMNATTIHN 604

Db 236 VIPSKEVLE----- 245

QY 605 LTFQNTNDEEFKMQITVEVDTREGPKLNSTAQGYENLVSPITLLPEAEITLFEDIPKEK 664

Db 246 -----VQP-AAVQSRV-----HR 258

QY 665 FPK-----FKRHDVNSTRAQAEVETIPLVNIISLLPKDAQLSLNTLDLQLEHGDITLKY 718

Db 259 FPQMGLOKLFRRSSANF-----KD-----VNRHR 282

QY 719 NLSKALLRFLMNSQAHAKIKNOAIITDETNDLSVAPQEKVHKSLPNSLGSYERLQRL 778

Db 283 NVSTKLRLRIVERNFNKAALMS---LNPELETSSSEPQTQTH----- 322

QY 779 TPAVSVKVNHDGQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIVPLESQMTK 838

Db 323 ----- 322

QY 839 EKKITGKEKENSMEENAEHIGVTEVLLGRKLOHYTDSYLGFLPWEKKYFLDLDEEE 898

Db 323 -----GLRKE----- 327

QY 899 SLKTQLAYFTDSKNRARKDRTADSLRYVYNKILNSKFGFTSRKVPFAHMHMIDRIVMQE 958

Db 328 -----DFKS-----STDIYSHSLIATNMLNRAIFKARHVLAHVGLDKDIVEA 373

QY 959 LQDMFPEEDKTSFKVHRSEDMQFAFSFYILMSAVQPLNISQVDEVDQSGVLSDR 1018

Db 374 MORFHOILDTAQRPAFTDLOYAFAYYSLMSETKVMSVEIFEEDFTDGSATWSDR 433

QY 1019 EIRTLATRIHELPSLODLTGLEHMLINCSK---MLPADITQLNNIPPTQESYDNLPL 1074

Db 434 EVRTFLTRIYOPPLDWSAMRYFEVWQCNTRNLGMHLKVDTVESHSL--VYERYEDSNLP 491

QY 1075 PVTKSLVYNCKPVTDKTHKAYKDKNRYRPELGE--EEIAFKMLRNVSHVVCQLDIRK 1132

Db 492 TITRDLVRCPLTAEALAAFAVAPKPNFHVSPKRTSHSNFMMLTSLNLTVEVSESLRLRR 551

QY 1133 NPKRFVCLNDNIDNH--KDAQTKAVLRDFEYEMFPISQFELPREYRNFRLHMHLEQW 1191

Db 552 NPKRFNCINDLNDANGEDNEMVPHLEDFYLSFFPPRSKFEPLPPQYRNFESWRDQFQW 611

QY 1192 R 1192

Db 612 K 612

RESULT 6

Q8SX14

ID Q8SX14 PRELIMINARY; PRT; 666 AA.

AC Q8SX14;

DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE RE35033p.

GN CG8027.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA George R., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,

RA Miranda A., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,

RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,

RA Celniker S.

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY089618; AAL50356.1; - F6PDB6D1C39248 CRC64;

SQ SEQUENCE 666 AA; 77745 MW; F6PDB6D1C39248 CRC64;

Query Match 13.5%; Score 854.5; DB 5; Length 666;

Best Local Similarity 24.4%; Pred. No. 6.9e-43;

Matches 220; Conservative 107; Mismatches 201; Indels 373; Gaps 15;

QY 305 DEDISASRFEDNEELRYSLRSIERHAPWVRNIFIVTNGQIPSWLNDNPRVTIVTHQDVF 364

Db 86 DDKYDPSRFDKNEELRYSLRSLEKHAWIIRVIVTNGQIPSWLDSYERVTVPVHEVLA 145

QY 365 RNLSHLPTFSSPAIESHVRIEGLSQFIYLNDDVFMFGKVDWPDYSHSKGQKVIYLTWP 424

Db 146 PDPQLPTFSSSAIETFLHRIPKSLRFLYLNDDIFLGAPLYPEDLYTEAGVRVYQAWM 205

QY 425 VPCNAEGCPGSKWIDGYCDKACNNSACDWDGDCSGNSGGSRYIAGGGTSGIGVQWPQ 484

Db 206 VPGCALDCPWTYIGDGACDRHCNIDACQFDGDCSETGPAS----- 246

QY 485 FGGGINSVSYCNOGCANSWLADKFCQACNVLSGDFDAGDGHFELYKVILLPNQTH 544

Db 247 -----DAH 249

QY 545 YIIPKGECLPYFSPAFAVAKRGVEGAYSDNPIIRHASIANKWKTIHLMHSGMNATTIHN 604

Db 250 VIPSKEVLE----- 259

QY 605 LTFQNTNDEEFKMQITVEVDTREGPKLNSTAQGYENLVSPITLLPEAEITLFEDIPKEK 664

Db 260 -----VQP-AAVQSRV-----HR 272

QY 665 FPK-----FKRHDVNSTRAQAEVETIPLVNIISLLPKDAQLSLNTLDLQLEHGDITLKY 718

Db 273 FPQMGLOKLFRRSSANF-----KD-----VNRHR 296

QY 719 NLSKALLRFLMNSQAHAKIKNOAIITDETNDLSVAPQEKVHKSLPNSLGSYERLQRL 778

Db 297 NVSTKLRLRIVERNFNKAALMS---LNPELETSSSEPQTQTH----- 336

QY 779 TPAVSVKVNHDGQGNPPLDLETTARFRVETHQKTIGGNVTKEKPPSLIVPLESQMTK 838

Db 337 ----- 336

QY 839 EKKITGKEKENSMEENAEHIGVTEVLLGRKLOHYTDSYLGFLPWEKKYFLDLDEEE 898



```

Db 337 -----GLRKE----- 341
QY 899 SUKTLQAYTDSKNRKYKRTFADSLRVVVKILSKGFTSRKYPAPAHMHMIDRVMOE 958
Db 342 -----DFKS-----STDIYSHSLIATNMLLNFRAYGFRARHVLAHVGFLLDKRDIVEA 387
QY 959 LODMPPEFDKTSFHKVHSEDMOFAFVYVYLMISAVQPLNISQVDFEVDVTDOSGVLSDR 1018
Db 388 MORRHOQILDTAHOFRAPTDLOQAFAYISFLMSETKVMSEVEEFDFFDGSATWSDR 447
QY 1019 EIRTLATRIHELPSLQDITGLEHMLINCSK----MLPADITQLNIPPTQESYDNP 1074
Db 448 EVRTFLTRIYQPPLDWSAMRYEEVVQCNTRNIGMHLKVDIVESHSTL--VVEREDSNLP 505
QY 1075 PYTKSLVTNCKPVTDIKHKAYKDKKRYFEINGE--EEIAFKMIRTNVSHVVQGLDDIRK 1132
Db 506 TITRDVWRCPLLAEALANFAVRPKYFNHVSFKRTSHSNFMMLTSLNTEVVESLDRLR 565
QY 1133 NPKRFVCLNDNDHNDH-KDAQTVKAVLRDEYVESMPPIPSQFELPREYRNRLHMHLEQEW 1191
Db 566 NPKFNCINDNDANRGEDNEVNRHLLDEFLYLFPPFRKSKFELPPQYRNRFESWRDFQW 625
QY 1192 R 1192
Db 626 K 626

RESULT 7
Q9NV34
ID Q9NV34 PRELIMINARY; PRT; 132 AA.
AC Q9NV34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CDNA FJ10959 f1s, clone PLACE1000562.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara K., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001821; BAA91926.1; -
SQ SEQUENCE 132 AA; 16083 MW; 92B5DEB4E062AD07 CRC64;

Query Match 7.38; Score 460; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2.6e-20;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1115 MIRTNVSHVVGOLDIRKNPKRFVCLNDNDHNDHDKAQTVKAVLRDFYESMFPPIPSQFEL 1174
Db 1 MIRTNVSHVVGOLDIRKNPKRFVCLNDNDHNDHDKAQTVKAVLRDFYESMFPPIPSQFEL 60
QY 1175 PREYRNRLHMHLEQEWYARDKILK 1199
Db 61 PREYRNRLHMHLEQEWYARDKILK 85

RESULT 8
Q9L1I2
ID Q9L1I2 PRELIMINARY; PRT; 602 AA.
AC Q9L1I2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

```

```

DE Hypothetical protein SCO2594.
GN SCO2594 OR SCC88.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Ouail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL139298; CAB75375.1; -
KW Hypothetical protein.
SQ SEQUENCE 602 AA; 67501 MW; 720EAB86A31D943F CRC64;

Query Match 5.28; Score 327; DB 16; Length 602;
Best Local Similarity 42.7%; Pred. No. 2.1e-11;
Matches 61; Conservative 33; Mismatches 45; Indels 4; Gaps 1;

QY 280 IDGKELTISPAYLLWDLISAISQKQEDDISARFEDNEELRYSLRSTIERHAPVNRITFV 339
Db 297 VDGSD----PAWINSRAEFSRDPYHEEAANAARYLSRDELRYSLNLYAPVWRNLYLV 352
QY 340 TNGQIPSWLNDMPRTIVTHQDVFRNLHLPTFSSPAIESHVHIEGLSKQFIYLNDDV 399
Db 353 TADQTPDLNTHDPRKLVVSHKKEIFSEPTSLPTFNSHAIESQLHHIDGLSEHLYFNDDV 412
QY 400 MFGKDVMPDDFYSHSGKQKVLVT 422
Db 413 MLGRETLPOHQHFFLPNGLGQYLS 435

RESULT 9
Q9L1I4
ID Q9L1I4 PRELIMINARY; PRT; 942 AA.
AC Q9L1I4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transferase.
GN SCO2592 OR SCC88.03C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

```

```

RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL139298; CAB75373.1; -.
DR InterPro: IPR001296; Glycos_transf_1.
DR Pfam: PF00534; Glycos_transf_1; 1.
DR Transferase.
SQ SEQUENCE 942 AA; 105063 MW; 108B737B98A97B80 CRC64;

Query Match 5.1%; Score 324.5; DB 16; Length 942;
Best Local Similarity 24.9%; Pred. No. 5.6e-11;
Matches 101; Conservative 75; Mismatches 156; Indels 73; Gaps 13;

QY 93 EQKAMREILGKNTTEPTKKSEKQLECLLTHCIKVPMLVLDLPALPANITLKDLPSPSEFH 152
DB 428 DQAREVVIGASDRSLVRAGRLSE-----VRDDLQSGSEIVQRFETVVEALE 475

QY 153 SASDFINVAKPK--NPSTNVSVVFDSTKDVDAHSGLLKGNRSQTVMRGYLTDTKEVPG 210
DB 476 SSGIPYVLLRDDNPNRRRLAVDAAEQTR-VRKALAGAYEG---KAVYAECLKPRTHAPG 531

QY 211 LVLMQ-----DLAFLSGFPPTKFNQLK-----TKLPENLSK----- 244
DB 532 VLLAERLEAVEGAVGLRFRPVVTSTLRFPGAYGCDIEFWQVPEEGGQGFVAPLR 591

QY 245 -----VKLLQLYSPASVALLKLNPKDFQELNKQTKK-----NNTIDGKELTISPAYLLW 294
DB 592 PSAVGPKLPSTPDRTRV---KDREYPTLEPLTRKLVSDITFPDVAVTVWDDSDPRW 647

QY 295 D-----LSAISQSKQDEISASRFEDNEELRYSLRSIERHAPWVRNIFVTNGQIP 345
DB 648 QERRARRRAALGLEASGDE---AARFRNDELRYSLRSIAMFAPWIRKIYLVTDQTP 704

QY 346 SWNLNDPRVTIVTHODVFNRLSHLPTFSSPAIESHVHRIEGLSOKFIYLNDDVMFGKV 405
DB 705 EWLATEHEGIEVSHRDIFTDQDCLPTFNSHSIESQLHHIDGLSEQFLYLNDDVFIGRPV 764

QY 406 WPDFFYSHSKGQKYLTPWPVPCNAEGCPGSGWIKDGYCDKACNNSA 450
DB 765 GAQRFELPNGASRF--WSPTTVPGEP--TEDEGYFAAKNNRA 806

RESULT 10
ID O69853 PRELIMINARY; PRT; 586 AA.
AC O69853;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein SCO6023.
GN SCO6023 OR SC1C3.11.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL139298; CAB75373.1; -.
DR InterPro: IPR001296; Glycos_transf_1.
DR Pfam: PF00534; Glycos_transf_1; 1.
DR Transferase.
SQ SEQUENCE 942 AA; 105063 MW; 108B737B98A97B80 CRC64;

Query Match 5.1%; Score 321.5; DB 16; Length 586;
Best Local Similarity 46.3%; Pred. No. 4.3e-11;
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;

QY 280 IDGKELTISPAYLLW--DLSAISQSKQDEISASRFEDNEELRYSLRSIERHAPWVRNIF 337
DB 282 VDGND-----PA---WRRRRSAYDGGYHAESAARAARYISRDELRYSLRALEQNAQWVRHVH 334

QY 338 IVTNGOIPSWNLNDPRVTIVTHODVFNRLSHLPTFSSPAIESHVHRIEGLSOKFIYLNDD 397
DB 335 LVTDGQRPAWLNDSPRLTVVDHSEIFADPALPTFNSHAIESRLHKLHGLSEHFLYLNDD 394

QY 398 DVMFGKDVMPDDFY 411
DB 395 DMFLGRPTPQDF 408

RESULT 11
ID O69851 PRELIMINARY; PRT; 541 AA.
AC O69851;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein SCO6021.
GN SCO6021 OR SC1C3.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL023702; CAA19233.1; -.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 541 AA; 60084 MW; 45E93981D16C6D2 CRC64;

Query Match 4.9%; Score 313; DB 16; Length 541;
Best Local Similarity 39.2%; Pred. No. 1.2e-10;
Matches 67; Conservative 31; Mismatches 63; Indels 10; Gaps 5;

QY 280 IDGKELTISPAYLLWDLISAISQSKQDEISASRFEDNEELRYSLRSIERHAPWVRNIF 339
DB 242 VDGND-----PAWKQKQAQKGEVYHAESAARAFISRDELRYSLRHLFAPWIRNIYV 297

QY 340 TNGOIPSWNLNDPRVTIVTHODVFNRLSHLPTFSSPAIESHVHRIEGLSOKFIYLNDDV 399
DB 340 TNGOIPSWNLNDPRVTIVTHODVFNRLSHLPTFSSPAIESHVHRIEGLSOKFIYLNDDV 399

```

Db 298 TDDQVPAWREDLPQARIATHREIFRPNEDLPTFNHSHSIESQLHHIEGLAEHFLYFNDDM 357  
 QY 400 MFGKDVWDDDFVSHKQKQVLTWPVPN-CAEGCPGSGWIKGYCDKACNNS 449  
 Db 358 FMGRPVAPHSFPT-PNGTARY--FPSRNRIPOGVAE--TDSPPVDAACKNN 403

RESULT 12  
 Q9EVX1 PRELIMINARY; PRT; 238 AA.  
 AC Q9EVX1;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Putative hexose transferase.  
 GN CPSO.  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1304;  
 RN [1]  
 RX MEDLINE=20519245; PubMed=11065358;  
 RA Almiron-Roig E., Mulholland F., Gasson M.J., Griffin A.M.;  
 RT "The complete cps gene cluster from Streptococcus thermophilus NC9B  
 RT 2393 involved in the biosynthesis of a new exopolysaccharide.";  
 RL Microbiology 146:2793-2802(2000).  
 DR EMBL; Y17900; CAC18360.1; -.  
 KW Transferase.  
 SQ SEQUENCE 238 AA; 29110 MW; A964250B01B78988 CRC64;

Query Match 4.5%; Score 286; DB 2; Length 238;  
 Best Local Similarity 37.7%; Pred. No. 1.5e-09;  
 Matches 58; Conservative 28; Mismatches 44; Indels 24; Gaps 3;

QY 262 NPKDFQELNKKTKNMTIDGKELTISPAYLLWDLSAISQSKODEDISASRDEENELRY 321  
 Db 11 NDPEFREKNKYTPNHRKID-----NDED-NVHRYRDYGTFFY 47  
 QY 322 SLRSIERHAPWENIIVTNGQIPSWLNDNPRVITVTHQDVFNLSHLPTSSPAIESH 381  
 Db 48 WFRKVERHAPWNIYLTNGRPKWLNVNHPKLVKWRHEE-FIPKEYLPFIENSAIEMN 106  
 QY 382 VHRIGLSQFIYLNDDVFGKDVWPDFFYSHSK 415  
 Db 107 IHRIDGLSENFVLFNDMDYLIQDYKYSDFEVNEK 140

RESULT 13  
 Q9JWW8 PRELIMINARY; PRT; 545 AA.  
 AC Q9JWW8;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Putative capsule biosynthesis protein.  
 GN SABC OR NMA0200.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RX MEDLINE=2022556; PubMed=10761919;  
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,  
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,  
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 RA Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 RT meningitidis Z2491.";  
 RL Nature 404:502-506(2000).

DR EMBL; AL162752; CAB83514.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 545 AA; 64133 MW; 7D2564286C95CD3D CRC64;

Query Match 4.2%; Score 265; DB 16; Length 545;  
 Best Local Similarity 27.8%; Pred. No. 9.2e-08;  
 Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13;

QY 135 LPANITLKDLPSPYFSFHSASDIFNVAKPNKPNSTNVVVVFDSTKQVEDAHSGLLKG-NS 193  
 Db 97 LPSNLTGK--PALCILESHKEDFLN-----KFLTISSENKLYQKFGQIKNPKS 145  
 QY 194 RQTVWRGY-----LTTQKEVPGV-----LMQDLAFLSGFPPTFKETNQLKT 235  
 Db 146 VNEIWTDLFSIAHVDKMLSTDRTLSSISQFWRLEFCKEDKDFI-----LFTANRYSR 200  
 QY 236 KLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNMTIDGKELTISPAYLLWD 295  
 Db 201 KLWKH---SIKNQLFKEG-----IRNYSEISSLPYEEDHNFIDILVFTWVNSEDKNWQ 251  
 QY 296 LSASISQSKQD---EDISASRFEDNEELRYSLRSTIERHAPWVRNIFIVTNGQIPSWLNDN 352  
 Db 252 -ELYKKYKPDENSATSTSRFLSRDELKFAIRSWEMNGSFIRKIFIVSNCAAPPAILDLNN 310  
 QY 353 PRVTIVTHQDVFNLSHLPTSSPAIESHVRHIEGLSQFIYLNDDVFMFGKDVWPDFFYS 412  
 Db 311 PKIQWVHEEIMPO- SALPTFSSHAETSLHHIPGISNYFIYSNDDFLTKPLKNKNFFY 369  
 QY 413 HSKGQKVVLT-TWVPVN--CAEGCP 433  
 Db 370 SNGIAKLRLAEWNGNVECTEGEP 393

RESULT 14  
 O68215 PRELIMINARY; PRT; 545 AA.  
 AC O68215;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE SABC.  
 GN SABC.  
 OS Neisseria meningitidis (serogroup A).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=65699;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F8229;  
 RX MEDLINE=98175678; PubMed=9515923;  
 RA Swartley J.S., Liu L.-J., Miller Y.K., Martin L.E., Edupuganti S.,  
 RA Stephens D.S.;  
 RT "Characterization of the gene cassette required for biosynthesis of  
 RT the (alpha1-26)-linked N-acetyl-D-mannosamine-1-phosphate capsule of  
 RT serogroup A Neisseria meningitidis.";  
 RL J. Bacteriol. 180:1533-1539(1998).  
 DR EMBL; AF019760; AAC38286.1; -.  
 SQ SEQUENCE 545 AA; 64097 MW; 1BC3400C5CC33DCD CRC64;

Query Match 4.2%; Score 264; DB 2; Length 545;  
 Best Local Similarity 27.8%; Pred. No. 1.1e-07;  
 Matches 90; Conservative 52; Mismatches 130; Indels 52; Gaps 13;

QY 135 LPANITLKDLPSPYFSFHSASDIFNVAKPNKPNSTNVVVVFDSTKQVEDAHSGLLKG-NS 193  
 Db 97 LPSNLTGK--PALCILESHKEDFLN-----KFLTISSENKLYQKFGQIKNPKS 145  
 QY 194 RQTVWRGY-----LTTQKEVPGV-----LMQDLAFLSGFPPTFKETNQLKT 235  
 Db 146 VNEIWTDLFSIAHVDKMLSTDRTLSSISQFWRLEFCKEDKDFI-----LFTANRYSR 200  
 QY 236 KLPENLSSKVKLLQLYSEASVALLKNNPKDFQELNKKTKNMTIDGKELTISPAYLLWD 295  
 Db 201 KLWKH---SIKNQLFKEG-----IRNYSEISSLPYEEDHNFIDILVFTWVNSEDKNWQ 251

1

Search completed: April 13, 2003, 03:22:39  
Job time : 69.7872 secs



hypothetical protein DKFp762B226.1 - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: T50618  
R:Bloeker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, June 2000  
A:Reference number: 225143  
A:Accession: T50618  
A:Molecule type: mRNA  
A:Status: preliminary  
A:Residues: 1-248 <AAA>  
A:Cross-references: EMBL:AL359588  
A:Experimental source: adult melanoma (MeWo cell line); clone DKFp762B226  
C:Genetics:  
A:Note: DKFp762B226.1

Query Match 16.5%; Score 1046; DB 2; Length 248;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52;  
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1003 VFDEVDTSQGLSDREITLARIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIP 1062  
Db 5 VFDEVDTSQGLSDREITLARIHELPLSLQDLTGLEHMLINCSKMLPADITQLNNIP 64  
QY 1063 PTQESYDNPPLPPVTKSLVTCKPVTDKIKHAYDKNKYRPEIMGEIEAFKMTIRTVNSH 1122  
Db 65 PTQESYDNPPLPPVTKSLVTCKPVTDKIKHAYDKNKYRPEIMGEIEAFKMTIRTVNSH 124  
QY 1123 VVGQDDIRKPRKFCVCLNDINHNHDAQTIVKAVLRDFYESMPPIPSQFELPREYNRF 1182  
Db 125 VVGQDDIRKPRKFCVCLNDINHNHDAQTIVKAVLRDFYESMPPIPSQFELPREYNRF 184  
QY 1183 LHMHELQWRAIRYDKLK 1199  
Db 185 LHMHELQWRAIRYDKLK 201

## RESULT 3

T34703  
hypothetical protein SC1C3.11 SC1C3.11 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T34703  
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z21554  
A:Accession: T34703  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-586 <OLI>  
A:Cross-references: EMBL:AL023702; PIDN:CAAL9235.1; GSPDB:GN00070; SCOEDB:SC1C3.11  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC1C3.11

Query Match 5.1%; Score 321.5; DB 2; Length 586;  
Best Local Similarity 46.3%; Pred. No. 1.2e-10;  
Matches 62; Conservative 30; Mismatches 33; Indels 9; Gaps 3;  
QY 280 IDGKELTISPAYLLW--DLASAISQSKODEISASREFDNEELRYSLRSIERHAPWVRNIF 337  
Db 282 VDGND---PA---WRRRRAYGGVHAESAARVYSIRDELRYSLRALEFONAPWVRVH 334  
QY 338 IVTNGOIPSWNLNDNPRVTIVTHQDFRNLSHLPTFSSPAIESHVHRIEGLSOKFIYLLND 397  
Db 335 LVTDGQRPALWNSHPLTVVDHSEIFADPAALPTNSHAIESRLHIKGLSEHFYLLND 394  
QY 398 DVMEGKDVWDDFF 411  
Db 395 DMFLGRPVTPQDFF 408

## RESULT 4

Db 395 DMFLGRPVTPQDFF 408

## T34701

hypothetical protein SC1C3.09 SC1C3.09 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T34701  
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z21554  
A:Accession: T34701  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-541 <OLI>  
A:Cross-references: EMBL:AL023702; PIDN:CAAL19233.1; GSPDB:GN00070; SCOEDB:SC1C3.09  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC1C3.09

Query Match 4.9%; Score 313; DB 2; Length 541;  
Best Local Similarity 39.2%; Pred. No. 3.4e-10;  
Matches 67; Conservative 31; Mismatches 63; Indels 10; Gaps 5;  
QY 280 IDGKELTISPAYLLWDLASAISQSKODEISASREFDNEELRYSLRSIERHAPWVRNIFIV 339  
Db 242 VDGND---PAMKQKAAQKGEVYHAESAASDARFISRDRLYSIRSLHLFAPWIRNIYV 297  
QY 340 TNGOIPSWNLNDNPRVTIVTHQDFRNLSHLPTFSSPAIESHVHRIEGLSOKFIYLLNDV 399  
Db 298 TDDQVPAWRREDLPAGARIATHREIFNPEDLPTFNSHSIESQLHHIEGLAEHFLFNDDM 357  
QY 400 MFGKQVWPDFFSHSGQKQVYLTWPVN-CABGCPGSIKDGCDKACNNS 449  
Db 358 FMGRPVAPHSEFT-PNGTARY--FPSRRNIPQCAVAE--TDSVPDAAACKNN 403

## RESULT 5

CB2014  
probable capsule biosynthesis protein NMA0200 [imported] - Neisseria meningitidis (st  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: CB2014  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: CB2014  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-545 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83514.1; PID:g737  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: sacB; NMA0200

Query Match 4.2%; Score 265; DB 2; Length 545;  
Best Local Similarity 27.8%; Pred. No. 1.8e-07;  
Matches 90; Conservative 53; Mismatches 129; Indels 52; Gaps 13;  
QY 135 LPANITLKDLSLYSFHSADIFNVAKPNPSTNVSVVDFDSTKVDHAHSGLLKG-NS 193  
Db 97 LPSNLTLL--PALCILESHKEDFLN-----KFLLTISSENLLKQYKFGQIKNPKS 145  
QY 194 RQTVVRGY-----LTTDKVEPGLV-----LMQDLAFSLGPPPTFKETNQLKT 235  
Db 146 VNEIWTDLFSTAHDVDMKLTSTRTSSISQFWRLFECKEDKDFI-----LFPANRYSR 200  
QY 236 KLPENLSKVKLLQLYSEASVALLKNNPKDFQELNKTQKNMTIDGKELTISPAYLLWD 295  
Db 201 KLWKH---SIKNQLFKEG-----IRNYSEISSLPYEDHNFIDLVFTWVNSDKNQ 251  
QY 296 LSAISQSKOD---EDISASREFDNEELRYSLRSIERHAPWVRNIFIVTNGOIPSWNLND 352  
Db 252 -ELYKKYKPDFNSDATSTSFRLSRDELKFAIWSWMNGSFIRKTFIVSNCAAPPALWLLNN 310







QY 946 HMPHMDIRVMOELDMFPEEDFKTSFHKVRHSEDMQFAFSFYIYLMGAVQPLNISQVDF 1005  
Db 1274 -----LQKIIIEENDMITEFVEKQISEGVDSIEKQLPLSNLRYDFMEKREVNLIT---E 1325  
QY 1006 EVDTD-OSGVLSDREIRTLATRIHELPLSLDQLTGLEHMLNCSKMLPADITQLNNIPT 1064  
Db 1326 EVKTEENENNEEL-AYDRENKILLWL-----LYOQGVVQIKNI-HI 1368  
QY 1065 QESYDPLNPVTKSLVT-----NCKPVTDKIHKAYKDK 1098  
Db 1369 LTPYMPPEAYNFYKATTPFENVEYRCRLIVDAIAELYAKK 1409

## RESULT 11

S21801

myosin heavy chain, neuronal [similarity] - rat

N:Alternate names: myosin II

N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002

C:Accession: S21801; PN0013; S18134

R:Sun, W.; Chantler, P.D.

J. Mol. Biol. 224, 1185-1193, 1992

A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain

A:Reference number: S21801; MUID:92235856; PMID:1569576

A:Accession: S21801

A:Molecule type: mRNA

A:Residues: 1-1999 &lt;SUN&gt;

A:Cross-references: EMBL:X62659

R:Sun, W.; Chantler, P.D.

Biochem. Biophys. Res. Commun. 175, 244-249, 1991

A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral

A:Reference number: PN0013; MUID:91151356; PMID:1998509

A:Accession: PN0013

A:Molecule type: mRNA

A:Residues: 1914-1998, 'I' &lt;S2U&gt;

A:Experimental source: brain

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide

F:84-763/Domain: myosin motor domain homology &lt;NMOT&gt;

F:174-181/Region: nucleotide-binding motif A (P-loop)

F:341-575/Region: actin binding #status predicted

F:653-675/Region: actin binding #status predicted

F:836-1999/Domain: coiled coil #status predicted &lt;COI&gt;

F:836-1276/Region: S2

F:1277-1999/Region: light meromyosin

F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F:180/Binding site: ATP (Lys) #status predicted

F:693-703/Active site: Cys #status predicted

F:1916/Binding site: phosphate (Ser) (covalent) #status predicted

F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match

Best Local Similarity 2.8%; Score 176; DB 1; Length 1999;

Matches 136; Conservative 104; Mismatches 232; Indels 158; Gaps 28;

QY 612 DEEFKQITVEVDTRGGKLNSTAKGY-----ENLYSPITILLPEAEILFEDIPK 661

Db 850 DEELMKK--VEKQTKVEALEMERKHQOLLEKNILAEQLQAEITLFAEAEMRLAA 907

QY 662 EKRPFRKRDVNSTRRQAEKPIPLVNLISLLPKDAQLSLNLTDLQLEHGDITLKGYNLS 721

Db 908 KQELSEILLDLSESVEEERNOILQNEK---KKEQGHKNLDELQLEDMESARQKLE 964

QY 722 K-----SFLMNSQAKI-KNOAITD---ETNDSLVAPQEKQVHKSLPNS- 768

Db 965 KVTTEAKKLKEBEQIILEDQNKLAKEKKLLEDRTAEFTTNLTBEEESKSLAKLNKH 1024

QY 769 ----LGVSELQRLTPPAVSKVNGHQONPPLDLETTARFVETHQKTIGGNVTEK 824

Db 1025 EAMITDLEELRR-----EKKQOELE-----KTRRKLEGSDTSL 1060

QY 825 PPSLIVPLESOMTEKKITKEKEN-----SRMBEN-AENHIGVTEVLLGRKLQHYTDSY 878  
Db 1061 --DOIAELOAQIAELKQOLAKKEELOALARVEEAAQNMALKKI---RELESQISEL 1115  
QY 879 LGFLPWK-----KKYFLDLLDEESLKTQLAYFTDS-----KNRARKYKRTFADSLRYVN 929  
Db 1116 QEDLESERASRNKAQKQKRDGLGEELEALKTELEDLTDTAAQOELRSKRE-----QEVN 1169  
QY 930 KILNSKGFCTSRKVPAPHPHMDR--IVMOELQDMFPE----- 965  
Db 1170 -ILAKTLEEEAKTHEAQIQEMRQKHSQAVEELAEQLQEQTKRKKVANLEKAKQTLNENGEL 1228  
QY 966 -----EFDKTSFHKVRHSEDMQFAFSFYIYLMGAVQPLNISQVDFDVTDOGSVLSD 1017  
Db 1229 ANEVKVLLOGGRDSEHKRKVE-----AQLEQLOVK--FNE-----GE 1264  
QY 1018 REIRTLATRIHELPLSLDQLTGLEHMLNCSKMLPADITQL--NNIPTQESYDNPPLPPV 1076  
Db 1265 RVVTEADKVTKLQVELDNVTGLLSQSDSKSLTKDFSALESQLODTQELLOEENRQKL 1324  
QY 1077 TKSIVTNCKPVTDKIHKAYKDKNKYRFEIMGEERIEAFKMIRTNVSHVVGOLDIDIRKNPRK 1136  
Db 1325 --SLSTKLKQVED-----EKNSFREQLBEEEEEKHNLEKQIATLHAQVADMKKKMED 1375  
QY 1137 FV-CINDNIDHNHDKDAQTVKAVLRDFEYSM 1165  
Db 1376 SVGCL-----ETAEEVKRKLOKLEGL 1397

## RESULT 12

A26868

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jun-2000

C:Accession: A26868

R:Tanabe, K.; Mackay, M.; Goman, M.; Scalfe, J.G.

J. Mol. Biol. 195, 273-287, 1987

A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium

A:Reference number: A26868; MUID:88011243; PMID:3079521

A:Accession: A26868

A:Molecule type: DNA

A:Residues: 1-1701 &lt;TAN&gt;

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

F:1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F:20-1701/Product: major merozoite surface antigen #status predicted &lt;MAT&gt;

Query Match

Best Local Similarity 2.7%; Score 171; DB 2; Length 1701;

Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;

QY 107 EPTKKSEKQLECLLT--HCIKVPMVLDPALPANITLKDLPSPYSPHSASDIFNVAKPK 164

Db 27 ELVKKLEALEDAVLTYGSLFQKEKVMVLNCTSGTAVTTSTPGSSGCVSTGGSVASVASVA 86

QY 165 NPSTNVSVVVFSDTKDVEDAHSGLLKGNROT-----VWRGYLTDD 205

Db 87 SGGSGSV-----ASGG--SGNSRRTNPSDSSNTKTYADLKHRYQVNYLFTI 133

QY 206 KEVPLGLVLMQDLAFSGPPPTFKETNQLKTKLPENLSKVKLLQLYSEASVALLKLNPK 265

Db 134 KELK-----YPELFDLTNHLMT--LSKNVGVFKYLDGYEINELLYKLNFFY 179

QY 266 DF--QELNK-----QTKKNMTIDGKELTISPAYLWLDLSAISQSKD-----ED----- 307

Db 180 DLLRAKLMDACANSYCQIPFNKLRANELDLVKKIVFGYRKPLDNIDKNYGMKMEIKKN 239

QY 308 -----ISASRFEDNEE-----LVSLSRISERHAPWRNFIPTVNGO 343

Db 240 KTTANINELLEGSKKIDQKNADNEEGKKLYQAQNLFIYKQLEAHNLISVLEKR 299

QY 344 IPSWLNLDNPRVTIVTHQDVPRNLSHLPTFSSP-----ATESHVHRIEGLSKQFIY 394

Db 300 IDTLKKNENIKKLEIDIDKIKTDAENPTTSGKPNLPENKKKEVEGHEEKIEA-KTIK 358  
QY 395 LNDVFMFGKDWMPDDFYSHSGQGVYLT-----WVPNCAEGCPGSKIWDG 440  
Db 359 FNIDSLF-TDPLEYLYLRENKAVDVPKSDPTKSVQIPKVPYPN-----404  
QY 441 YCDKACNNSACDWDGDCSGSGSRYIAGGGTSGISGVGPQWFGGGINSVSYNOGCA 500  
Db 405 -----GIVYPLPLTDIHNLSAADND--K 425  
QY 501 NSWLADKFCDOACNVLCGFDAGD-CGQDHFHLYKYVILLPNQTHYIIPKGECLPYFSFA 559  
Db 426 NSY-----GDLMPDPTKEKINEKIITDNKERKI-----FI 455  
QY 560 EVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTI---HFNLTQNTNDEE-- 614  
Db 456 NNKKOID---LEEKININHTKEQNK-KLEDEYKSKDYEELEKFEYEMKNNFNDKDV 511  
QY 615 ---FKMQITVEVD-TREGPKLNSTAQGYENLVSPITLLPEAEILFEDIPKEKRPK--F 668  
Db 512 DKIFSARYTYNVEKORYNNKFPSS-----NNSVYNVQKLKALSYLEDSLRKGISEKDF 566  
QY 669 KRHDVNST-----RRAQEEV-----KIPLVNIISLLPKDAQLSLNTLD---LQLEHGDIT 714  
Db 567 NHYTLTKGLEADIKKLTETIKSENKILEKFNKGLTHSANASLEVS DIVKLOQKVLLI 626  
QY 715 LKGYNLSKALLRSFLMNSOHAQIKNOAIITD-----ETNDSLVAPOEKVHKSILP 766  
Db 627 KKIEDLRKIEL---FLKNAQ---LKDSIHVPNIYKPNKPEPYLYLVLUKEVDKLEKFI 680  
QY 767 NSLGVSEY---LQRLTFPAVSVKNGHDQGNPPL-----DLETTARFRVETHTQKTIG 817  
Db 681 KYVDMKREQAVLSITQPLVAASETTEDGGHSTHTLSQSGETEVEETEVE--TEETVG 737  
QY 818 ---GNVTKEKPPSLIPLVESQMTKEKKITGKEKENSMBEENAHIGVTEVLLGRKLQHY- 874

RESULT 13  
A54498  
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is  
C:Species: Plasmodium falciparum  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jun-2000  
C:Accession: A54498

R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V  
Mol. Biochem. Parasitol. 27, 291-302, 1988  
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmo  
A:Reference number: A54498; MUID:88142999; PMID:2449612  
A:Accession: A54498  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1701 <PPT>  
A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413  
C:Superfamily: major merozoite surface antigen  
C:Keywords: surface antigen

Query Match 2.7%; Score 170; DB 2; Length 1701;  
Best Local Similarity 19.4%; Pred. No. 0.27;  
Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;

QY 107 EPTKXSEKOLECLT--HCIKVPMVLDPALPANITLKDPLSPYSPFSHASDIFNWKPK 164  
Db 27 ELVKLEALEDAVLGTYSLFQKEKMLVNEGTSCTAVTTSTPGSGSVTSGGSVASVA 86  
QY 165 NPTNVSVVVFEDTRKDEDAHSGLLKGNSROT-----VWRGYLTLD 205  
Db 87 SGGSGSV-----ASGG--SGNSRRTPNSDSDSNKTYADLKHVRQVYLF 133  
QY 206 KEYPGLVLMQDLAFLSGFPPTKETNQLTKLPENLSKVKLLQLYSEASVALLKNPK 265  
Db 134 KELK-----YPELFDLTHMLT-LSKNVDFGFKYLDGYYEINELLYKLFY 179  
QY 266 DF--QELNK-----QTKKNMTIDGKELTISPAYLLMDLSAISOSKOD-----ED----- 307  
Db 180 DLLRAKLNDAACANSYCOIPFNKIRANELDVLKIVFYGRPLDNKONVKMEDYIKKN 239  
QY 308 -----ISASREFNEE-----LRYSLRSIERHAPWVRNIFVTNGO 343  
Db 240 KTIANINELIGSKKTDONKNADNEEGKKLYOQINLFYNNKOLQEAHNLISVLEKR 299  
QY 344 IPSWLNLDNPRVTIYTHQDVFRNLSHLPTFSSP-----AIESHVHREGLSKRFY 394  
Db 300 IDTLKKNENIKKLEIDIDKIKTDAENPTTSGKPNLPENKKKEVEGHEEKIEA-KTIK 358  
QY 395 LNDVFMFGKDWMPDDFYSHSGQGVYLT-----WVPNCAEGCPGSKIWDG 440  
Db 359 FNIDSLF-TDPLEYLYLRENKAVDVPKSDPTKSVQIPKVPYPN-----404  
QY 441 YCDKACNNSACDWDGDCSGSGSRYIAGGGTSGISGVGPQWFGGGINSVSYNOGCA 500  
Db 405 -----GIVYPLPLTDIHNLSAADND--K 425  
QY 501 NSWLADKFCDOACNVLCGFDAGD-CGQDHFHLYKYVILLPNQTHYIIPKGECLPYFSFA 559  
Db 426 NSY-----GDLMPDPTKEKINEKIITDNKERKI-----FI 455  
QY 560 EVAKRGVEGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTI---HFNLTQNTNDEE-- 614  
Db 456 NNKKOID---LEEKININHTKEQNK-KLEDEYKSKDYEELEKFEYEMKNNFNDKDV 511  
QY 615 ---FKMQITVEVD-TREGPKLNSTAQGYENLVSPITLLPEAEILFEDIPKEKRPK--F 668  
Db 512 DKIFSARYTYNVEKORYNNKFPSS-----NNSVYNVQKLKALSYLEDSLRKGISEKDF 566  
QY 669 KRHDVNST-----RRAQEEV-----KIPLVNIISLLPKDAQLSLNTLD---LQLEHGDIT 714  
Db 567 NHYTLTKGLEADIKKLTETIKSENKILEKFNKGLTHSANASLEVS DIVKLOQKVLLI 626  
QY 715 LKGYNLSKALLRSFLMNSOHAQIKNOAIITD-----ETNDSLVAPOEKVHKSILP 766  
Db 627 KKIEDLRKIEL---FLKNAQ---LKDSIHVPNIYKPNKPEPYLYLVLUKEVDKLEKFI 680  
QY 767 NSLGVSEY---LQRLTFPAVSVKNGHDQGNPPL-----DLETTARFRVETHTQKTIG 817  
Db 681 KYVDMKREQAVLSITQPLVAASETTEDGGHSTHTLSQSGETEVEETEVE--TEETVG 737  
QY 818 ---GNVTKEKPPSLIPLVESQMTKEKKITGKEKENSMBEENAHIGVTEVLLGRKLQHY- 874

Db 738 HTTITITLPPK-----EESAPKEVKV-----ENSHKSDNSDQALTKTYLKLDEFL 788  
QY 875 TSYLGFLPWEKKYFL-----DLL-----DEESUKTQLAYFTDSKNRARKYRD 919  
Db 789 TKSII-----CHKYILVSNSSMDQKLELVNLTPEENELKS----- 825  
QY 920 TFADSLRVYKNTLSKFGFTSRKYPAHMPHMDRIVMLOELDMPEEPEDKTSFHKVRHSE 979  
Db 826 --CDPLDLFLNQN-----IPA--WYSLYDSMNI--DLQHLFFELFYK----- 863  
QY 980 DMQFAFSFYILMSAVQPLNTISQVDE-----VDTQSGVLSDREIR- 1021  
Db 864 -----EMYYLHLKEENHIKKLEEQKQITGTSTSPGNTVNTAQSATHSNQOQ 917  
QY 1022 -----TLATRIHELPLSL-----QDUTGLEHMLI--NCSKM-----LPAD 1054  
Db 918 SNASSTNTQGVAVSGPAVVEESH-PLTVLSISNDLKGIVSLNLGNKTKVPNPLTIS 976  
QY 1055 ITQL-----NNIPPTQESYDPNLPVPTKSLVNTCKPVT-----DKTHKAYDK- 1098  
Db 977 TTEMKEFYENILKNDITFNDIDKQFVKS-----NSKVTGLTETQKNALNDEI--KKLKDTL 1032  
QY 1099 -----NKYRPEI-----MGEETAFKMIKRTNVSHVVGQLDDIRKNPRKFCVCLND 1142  
Db 1033 QLSFDLYNKYKLDRLFNKKELGQDKQWKIKLLTLLKEQLESKLNSL--NPNHN--VLQNF 1090  
QY 1143 NIDHNH 1149  
Db 1091 SVFFNKK 1097

## RESULT 14

A49128

cell-fate determining gene Notch2 protein - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 02-Aug-2002

R:Weinmaster, G.; Roberts, V.J.; Lemke, G.

Development 116, 931-941, 1992

A:Title: Notch2: a second mammalian Notch gene.

A:Reference number: A49128; MUID:93202015; PMID:1295745

A:Accession: A49128

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-2471 &lt;WEI&gt;

A:Experimental source: Schwann cell

A&gt;Note: sequence extracted from NCBI backbone (NCBIP:127811)

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

F:264-295/Domain: EGF homology &lt;EGX1&gt;

F:799-830/Domain: EGF homology &lt;EGF1&gt;

F:877-908/Domain: EGF homology &lt;EGX2&gt;

F:1029-1060/Domain: EGF homology &lt;EGF&gt;

F:1067-1098/Domain: EGF homology &lt;EGX3&gt;

F:1153-1184/Domain: EGF homology &lt;EGF3&gt;

F:1191-1222/Domain: EGF homology &lt;EGX4&gt;

F:1876-1908/Domain: ankyrin repeat homology &lt;AN1&gt;

F:1909-1941/Domain: ankyrin repeat homology &lt;AN2&gt;

F:1943-1975/Domain: ankyrin repeat homology &lt;AN3&gt;

F:1976-2008/Domain: ankyrin repeat homology &lt;AN4&gt;

F:2009-2041/Domain: ankyrin repeat homology &lt;AN5&gt;

Query Match 2.7%; Score 169.5; DB 2; Length 2471;

Best Local Similarity 26.2%; Pred. No. 0.51;

Matches 44; Conservative 15; Mismatches 58; Indels 51; Gaps 5;

QY 407 PDFYSHSGQKYLWPNCAEGCPGSKWIKDYCDKACNNSACDWDGDCSGNSGGR 466

Db 1404 PPFWSHCESYTAPTSTPATCLSQYCADKARDGICDEACNSHACQWDGDC----- 1455

QY 467 YIAGGGGTSGVGPQW-QFGGGINSVSYCNOQC-----AN 501

Db 1456 -----SLTWEDFWANTSSLRCEWYINNOCDLNCNTAECLDFECQNSKCKYD 1506

QY 502 SWLADKF-----CDACNVLSGFDAGDGDHFEHYK-----VILLP 540  
Db 1507 KYCADHFKNHCDKGCNNECGWDGDLCDRAADQENLAEGILIVVLLP 1554

## RESULT 15

S06434

homeotic protein lin-12 precursor - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 20-Sep-1999

C:Accession: S06434; A24769

R:Yochem, J.; Weston, K.; Greenwald, I.

Nature 335, 547-550, 1988

A:Title: The Caenorhabditis elegans lin-12 gene encodes a transmembrane protein with

A:Reference number: S06434; MUID:88334747; PMID:3419531

A:Molecule type: DNA

A:Residues: 1-1429 &lt;YOC&gt;

A:Cross-references: EMBL:M12069; NID:gl56357; PIDN:AAA70191.1; PID:gl56358

R:Greenwald, I.

Cell 43, 583-590, 1985

A:Reference number: A24769; MUID:86079540; PMID:3000611

A:Accession: A24769

A:Molecule type: DNA

A:Residues: 173-712 &lt;GRE&gt;

C:Genetics:

C:Introns: 50/2; 90/1; 109/1; 172/3; 545/1; 589/2; 632/2; 1273/3; 1389/3

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

C:Keywords: glycoprotein; transmembrane protein

F:254-284/Domain: EGF homology &lt;EGF1&gt;

F:507-540/Domain: EGF homology &lt;EGF&gt;

F:547-578/Domain: EGF homology &lt;EGF2&gt;

F:909-931/Domain: transmembrane #status predicted &lt;TM&gt;

F:1093-1125/Domain: ankyrin repeat homology &lt;AN1&gt;

F:1206-1238/Domain: ankyrin repeat homology &lt;AN2&gt;

F:1240-1272/Domain: ankyrin repeat homology &lt;AN3&gt;

Query Match 2.6%; Score 166.5; DB 2; Length 1429;

Best Local Similarity 19.8%; Pred. No. 0.33;

Matches 111; Conservative 43; Mismatches 166; Indels 241; Gaps 21;

QY 428 CAEGCPGSKWIKDYCDKACNNSACDWDGDCSGNSGGRYIAGGGGTSGVGPQW--- 484  
Db 643 CSERA-----NDGNCDDADCNVACKFDGDCSGKR-----EPFSKCR 679

QY 485 -----FGGINSVSYCNOGCAN-----SWLADKFCD 510  
Db 680 YGNMCADFFANGV-----CQACNNECLYDGMDCLPVAVRCPVKIREHCASRFANGICD 734

QY 511 QACNVLSGFDAGDGDHFEHYKVLIPNQHYIIPKGECLPYFSFAEVAKRGVEGAY 570  
Db 735 PECTNGCGFDGDC-----DNET----- 753

QY 571 SDNPIIRHASTANKWKTHLIMHSGMNATTI--HENLTFQNTNDEPK----- 616  
Db 754 -----NATIITNRIIVQ--MDPFOVQTGGQSLMEISSA 786

QY 617 MQITVEVD--TREGPKLNSTAQGYENLVSPITL-----LPEAEILFEDIPEKRPKPKRH 671  
Db 787 LRVTVRIQRDEEGPLV---FQWNGESEMDRVKMNERQLTEQHVLSSTISR-----KIKRS 838

QY 672 DVNSTRRAQAEVKIPLVNIISILLPKDAQLSLNTLDLQLEHGDITLKGYNLS-----KS 723  
Db 839 ATNIGVVVYLEVQENCDTGKLYKDAQSVSDSISARLAKKAGIDSGFIPISALVAEPKRS 898

QY 724 ALLRSFL-----MNSQAHAKIN 740  
Db 899 GNTGTFLSNALLLIGACCLIVVVLGALPGNTRKRNRINASVMPMPNEKRNKN 958

QY 741 QAITDETNDLSLVAPOEKQV-----HKSILPNSLGVSE-----RLQRUTFP 781  
Db 959 HQSITSSQSHLSLEASYDGYIKQRNELQHSILYPNPGYNGNGDFLGNHTNLQIPTEP 1018

Search completed: April 13, 2003, 03:23:40  
Job time : 49.8703 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2003, 01:36:25 ; Search time 22.9235 Seconds  
(without alignments)  
2169.392 Million cell updates/sec

Title: US-10-023-888-2

Perfect score: 6340

Sequence: 1 METDTLLLVLLWLLVPGSTG.....NRFLHMHLEQEWRAVYDKLK 1199

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	171	2.7	1701	1 MSP1_PLAFM	P08569 plasmodium
2	170	2.7	1701	1 MSP1_PLAFM	P13819 plasmodium
3	169.5	2.7	2470	1 NTC2_MOUSE	Q35516 mus musculus
4	169.5	2.7	2471	1 NTC2_RAT	Q9QW30 rattus norv
5	166.5	2.6	1429	1 L112_CAEEL	P14585 caenorhabdi
6	166.5	2.6	2471	1 NTC2_HUMAN	Q04721 homo sapien
7	164	2.6	2531	1 NTC1_RAT	Q07008 rattus norv
8	163.5	2.6	2437	1 NTC1_BRARE	P46530 brachydania
9	163.5	2.6	3135	1 S230_PLAFO	Q08372 plasmodium
10	162.5	2.6	1005	1 RA50_METUA	Q58718 methanococc
11	162	2.6	2531	1 NTC1_MOUSE	Q01705 mus musculus
12	158.5	2.5	1961	1 MYH9_RAT	Q62812 rattus norv
13	158	2.5	1163	1 SBCC_CLOAB	Q971K1 clostridium
14	157.5	2.5	1978	1 MYHB_CHICK	P10587 gallus gall
15	155.5	2.5	2321	1 NTC3_HUMAN	Q9UM47 homo sapien
16	154.5	2.4	1251	1 RBP2_PLAVB	Q00799 plasmodium
17	154.5	2.4	1682	1 MSP1_PLAF3	P19598 plasmodium
18	153.5	2.4	1960	1 MYH9_HUMAN	P35579 homo sapien
19	153.5	2.4	2663	1 CENE_HUMAN	Q02224 homo sapien
20	153	2.4	1169	1 SMC_METJA	Q29037 methanococc
21	153	2.4	1225	1 SMC1_YEAST	P04934 plasmodium
22	153	2.4	1726	1 MSP1_PLAFC	P05495 plasmodium
23	153	2.4	1726	1 MSP1_PLAFC	P04945 plasmodium
24	153	2.4	1964	1 NTC4_MOUSE	P31695 mus musculus
25	152.5	2.4	2003	1 NTC4_MOUSE	Q99468 homo sapien
26	152	2.4	1957	1 YD86_SCHPO	Q10411 schizosacch
27	150.5	2.4	1526	1 MYSL2_SCHPO	Q9US16 schizosacch
28	150.5	2.4	1630	1 MSP1_PLAFK	P04932 plasmodium
29	150.5	2.4	1639	1 MSP1_PLAFW	P04933 plasmodium
30	149	2.4	2319	1 NTC3_RAT	Q9R172 rattus norv
31	148.5	2.3	2524	1 NTC3_XENLA	P21783 xenopus lae
32	148	2.3	2318	1 NTC3_MOUSE	Q61982 mus musculus
33	147	2.3	1031	1 RAD2_YEAST	P07276 saccharomyc

34	146	2.3	1875	1 MLP1_YEAST	Q02455 saccharomyc
35	145.5	2.3	2095	1 RPL_MOUSE	P56716 mus musculus
36	145.5	2.3	2556	1 NTC1_HUMAN	P46531 homo sapien
37	143	2.3	1818	1 HMW2_MYCPN	P75471 mycoplasma
38	141.5	2.2	1744	1 TANA_XENLA	Q01550 xenopus lae
39	140	2.2	3685	1 DMD_HUMAN	P11532 homo sapien
40	138	2.2	1928	1 MYSL_YEAST	P08964 saccharomyc
41	137.5	2.2	908	1 ATMB_SALTY	P22036 salmonella
42	137	2.2	1328	1 YMU0_YEAST	Q04670 saccharomyc
43	137	2.2	1939	1 MYH6_MESAU	P13539 mesocricetu
44	136.5	2.2	1295	1 GLP1_CAEEL	P13508 caenorhabdi
45	136.5	2.2	1372	1 FUS1_SCHPO	Q10719 schizosacch

## ALIGNMENTS

RESULT 1  
MSP1\_PLAFM  
ID MSP1\_PLAFM STANDARD; PRT; 1701 AA.  
AC P08569;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMMSA) (P190).  
GN MSP-1.  
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=70153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88011243; PubMed=3079521;  
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;  
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum."  
RL J. Mol. Biol. 195:273-287(1987).  
RN [2]  
RP REVISIONS TO 1403; 1569 AND 1629.  
RA Tanabe K.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-115 FROM N.A.  
RX MEDLINE=86136024; PubMed=3004972;  
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
RA Stunnenberg H., Buyard H.;  
RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level."  
RL EMBO J. 4:3823-3829(1985).  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (Potential).  
CC -!- PWM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X05624; CAA29112.1; -.  
DR PIR; A26868; A26868.  
DR PIR; B25120; B25120.  
DR InterPro; IPR000561; EGF-like.  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 2.7%; Score 171; DB 1; Length 1701;  
Best Local Similarity 19.4%; Pred. No. 0.057;  
Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;

QY 107 EPTKSEKQELCLT--HCIRVPMVLVDPALPANITLKDLPSPFHSASDIENVAKPK 164  
DB 27 ELVKKLEALEDAVLGTGYSLFQKEMVNLNDEGTSGTAVTTSTPGSSGSGVTSGGVSASVASVA 86

QY 165 NPSTNVSVVFDSTKDVDAISGLLKNSROT-----VWRGYLTID 205  
DB 87 SGGSGSV-----ASGG--SGNSRRTNPSDSDSNSTKYADLKHRVQNYLFTI 133

QY 206 KEVPLGLVMDLAFSLGPPPTFKETNOLTKLPENLSKVLLQLQYSEASVALLKLNPK 265  
DB 134 KELK-----YPELFDLTNHLT--LSKNVDGFKYLDIDYBEINELLYKLNYY 179

QY 266 DF--QELNK-----QTKNMTIDGKELTISPAYLLWDLISAISQSKD---ED----- 307  
DB 180 DLLRAKLANDACANSYCQIPFNKIRANELDVLKIVFGYKPLDNKDNVGMEDYIKKN 239

QY 308 -----ISASRFENEE-----LRSLSRERHAPVWRNIFVTNGQ 343  
DB 240 KTTIANINELIEGSKTTIDQKNADNEEGKKLYQAQYNLFYTNKQLQEAHNLISVLEKR 299

QY 344 IPSWLNLDNPRVTIVTHODVFNLSHLTFSSP-----AIESHVHRTLEGLSQKFIY 394  
DB 300 IDTLAKNENIKLLIEDIKYTDAPNTGSKPNLPENKKEVEGHEKKEIA-KTIK 358

QY 395 LNDVFMGKDVMPDFYSHSGKQKYLT-----WVPVNCACGCPGSKWIDK 440  
DB 359 FNIDSLF--TDPLEYLYLREKNKKVDVTPKSDPTKSVQIPKVPYPN----- 404

QY 441 YCDKACNSACDWDGDCSGSGSRYTAGGGTSGISGVGPQWQFGGINSVSYNCQCA 500  
DB 405 -----GIVYPLPLTDIHNLSAADND--K 425

QY 501 NSWLADKFCDAQCNVLSGCFDAGD--CGODHFHELYKVILLPNOTHYIIPKGBCLPYFSFA 559  
DB 426 NSY-----GDLNPDTEKINEKIITDNKERKI-----FI 455

QY 560 EVAKRGVEGAYSDNPPIIRHASIANKWKTHLIMHSGMNATTI---HFNLTQNTNDEE-- 614  
DB 456 NNKKOID--LEKNINHTKEQNK--KLEDEYKSKDVEELLEKEFYEMKFNPNFNDKDV 511

QY 615 ---FKMQITVEVD--TREGPKLNSTAKQGYENLVSPITLLPEAIILFEDIPKEKRPK--F 668  
DB 512 DKIFARSYTYNVEKORYNKKFSSS-----NNSVYNVQVKKKALSYLEDSLRKISERDF 566

QY 669 KHDVNSY-----BRAQEV-----KIPLVNIISLLPKDAQSLNTLD--LOLEHGDI 714  
DB 567 NHYITLTKGLEADIKKLEEKSSSENKILEKNFKGLTHSANASLEVSIDVLQVQKVLLI 626

QY 715 LKGYNLSALLRSFLMNSHAKIKNOAIITD-----ETNDSLVAPOEKQVHKSILP 766  
DB 627 KKIEDLRKIEL--FLKNAQ--LKDSITHVNIYKPNKPEPYLYILVKKEVDKLEKEFP 680

QY 767 NSLGVSE---LQRLTFPAYSVKVNGHDQGNPL-----DLETTARFVETHQKTI 817  
DB 681 KYKMDLKEQAVALSSITQPLVAASSETTDDGGHSTHTLSQSGTEVTEVTEV---TEETVG 737

QY 818 --GNVTKEKPPSLIPLVESQMTKEKIKTGKKEKNSRMEENAHNIGVTEVLGRKLQHY- 874  
DB 738 HTTIVTILPK-----EESAPKEVKV-----ENSIEHKSNDNSQALTKTYVLKLDLDEF 788

QY 875 TDSYLGFLPWEKKYFL-----DLL-----DEESLKTQLAYFTDSKNRARYKRD 919  
DB 789 TKSyi-----CHKYILVSNSSMDQKLELVNLTPEEKEKLS----- 825

QY 920 TFDASLRVYNKILNSKFGFTSRKYPAPHPMDRIDRVMLQDMPPPEEDKTSFHKVHSE 979  
DB 826 --CDPLDLFLIQQN-----IPA-MYSLYDS-MNNDLQHLFFELYOK----- 863

QY 980 DMQFAFSYFYILMSAVOPLNISQVDE-----VDTQSGVLSDEIR- 1021  
DB 864 -----EMYYLHLKKEENHIKKLEEQKIQGTSTSSPGNTVTYTAQSATHSNSQNO 917

QY 1022 -----TLATRIHELPLSL-----QDITGLEHMLI--NCSKM---LPAD 1054  
DB 918 SNASSTNTQNGVAVSSGPAVVEESH-PLTVLSISNDLKGIVSLNLGNKTKVPNPLAIS 976

QY 1055 ITOL-----NNIPPTQESYDNLPPVTKSLVTKCPVT-----DKIKKAYKDK- 1098  
DB 977 TTEMEKFEYENILKNDTYFNDDIKQFVKS---NSKVITGLTETQKNALNDEI-KKLDL 1032

QY 1099 -----NKYRFEI-----MGEETAFKMIKRTNVSHVVGQDLDIRKKNPKFVCLND 1142  
DB 1033 QLSFDLYNKYKLDRLFNKKELGQDKMQIKKLTLLKEQLESKLSNL-NNEPHN-VLQNF 1090

QY 1143 NIDHNH 1149  
DB 1091 SVFENK 1097

## RESULT 2

MSPI\_PLAFF ID MSPI\_PLAFF STANDARD; PRT; 1701 AA.  
AC P13819;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Merozoite surface protein 1 precursor (Merozoite surface antigens)  
DE (PMWSA).  
GN MSP-1.  
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5837;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88142999; PubMed=2449612;  
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,  
Brown G.V., Anders R.F., Kemp D.J.;  
RT "Variation in the precursor to the major merozoite surface antigens  
of Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 27:291-302(1988).  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
(Potential).  
CC -!- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42  
kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DB EMBL; M19143; AAA29653.1; -  
DB PIR; A54498; A54498.  
DB InterPro; IPR000561; EGF-like.  
DB Pfam; PF00008; EGF; 1.

KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1701 AA; 193719 MW; 3920875E73D38552 CRC64;  
Query Match 2.7%; Score 170; DB 1; Length 1701;  
Best Local Similarity 19.4%; Pred. No. 0.065;  
Matches 246; Conservative 164; Mismatches 437; Indels 420; Gaps 62;  
QY 107 EPTKKEKOLECLLT--HCIKVPMVLDPALPANITLKDLPSLYPSFHSASDIFNVAKPK 164  
DB 27 ELVKKLEALEDAVLGTGYSLFQKEKVMVNEGTSGETAVTTSTPGSSGVSVTGGSVASVA 86  
QY 165 NPSTNVSVVDFSTKDVEDAHSGLKNGSRQT-----ASGG--SGNSRRTNPSDSSNTKYADLKHVRQVYLFI 133  
DB 87 SGGSGGVS-----ASGG--SGNSRRTNPSDSSNTKYADLKHVRQVYLFI 133  
QY 206 KEVPLGLVLQDLAFLSGPPTKETNQLTKLPENLSSKVKLLQYSEASVALLKLNPK 265  
DB 134 KELK-----YPEFLDTNHMLT--LSKNVDGFKYLIDGYEINELLYKLNFFY 179  
QY 266 DF--QELNK-----OTKKNMTIDGKELTISPAYLLWLSAISOSKOD-----ED----- 307  
DB 180 DLLRAKLNACANSYCOIFENKIRANELDVLKKVFGYRKPLDNKIDNVGMEDYIKKN 239  
QY 308 -----ISARFEDNEE-----LRYSLRTERHAPWVRNFIWITNGQ 343  
DB 240 KTTIANINELIEGSKTTIDQKNADNEGKKLYQAQYNYFNKQLQEAHNLISVLEKR 299  
QY 344 IPSWLNDNPRVTIVHQDVFNLSHLPTFSP-----AISHVHRIEGLSOKTIY 394  
DB 300 IDTLKKNENIKLLEDIDIKRTDAENPTTGSKNPLPENKKEVSGHEEKIEIA--KTIK 358  
QY 395 LNDVVMFGKDVWPDFFYSHSGKQVYLT-----WVPNCAEGCPGSWIKDG 440  
DB 359 FNIDSLF--TDPLEYLLREKKNKVDVTPKSDPTKSVQIPKVPYPN----- 404  
QY 441 YCDKACNNSACDWGDCSGNSGSGRYTAGGGGTSGYGVGPWFQFGGINSVSYCNOGCA 500  
DB 405 -----GIVPLPLTDIHNLSLAADND--K 425  
QY 501 NSWLADKFCDAQNLVSCGFDAGD--CGQDFHELYKVVILLPNQTHYIIPKGECLPYFSA 559  
DB 426 NSY-----GDLNPDTEKEIKETITDNKERKI-----FI 455  
QY 560 EVAKRGVEGAYSDNPITRIASIANKWKTHLIMHSGMATTI---HFNLTFTQNTDDE-- 614  
DB 456 NNKKOID---LEERNINTKQNK--KLEDYEKSKDYEELKFEYEMKFNNDKDV 511  
QY 615 ----FRMQITVEVD--TREGPKLNSTAQGYENLVSPITLLPEAEILFEDIPKEKPPK--F 668  
DB 512 DKIFSARYTYNVEKQRYNNKFESS-----NNSVYNVQKLKALSYLEDSYLRKIGSEKDF 566  
QY 669 KRDVNST-----RRAQEV-----KPLVNISSLPKDAQLSNLTD---LQLEHGDI 714  
DB 567 NHYITLKTGLEADIKKLEIKESSENKILEKNFKGLTHSANASLEVSIDVQLQKVLLI 626  
QY 715 LKGYNLKSALLRSFLMNSOHAQIKNQAIITD-----ETNDSLVAPOEKQVHKSILP 766

DB 627 KKIEDLRKIEL---FLKNAQ---LKDSTHVPIIYKPNKPEYYLIIVLKEVDKLFEP 680  
QY 767 NSLGSYER---LORLTPEPAVSVKNGHDQGNPPL-----DLETTARFRVETHTQKTIG 817  
DB 681 KVKDKLKEQAVLSITQIPLVAASETTEDGGHSTHTLSQSGETEVTETEY---TEETVG 737  
QY 818 --GNVTKKPPSLIVPLESQMTKEKKTITGKEKNSRMEENHIGVTEVLLGRKLQHY- 874  
DB 738 HTTWTITLPPK-----EESAPKEVKV---ENSIEHKSNDNSOALTKTLYLKKLDEFL 788  
QY 875 TDSYLGFLPWKKYFL-----DLL-----DREESLKTOLAYFTDSKNRARIKRD 919  
DB 789 TKSYI-----CHKYILVSNSSMDOKLELVNLTPEENELKS----- 825  
QY 920 TFADSLRVNKLKSGFTSRKVPAPMHPMIDRIVMOELQDMPPEBFDKTSFHKVRHSE 979  
DB 826 --CDPLDLLFNQNN-----IPA-MYSLYDSMNI-DLQHLFFELYQK----- 863  
QY 980 DMQFAFSFYIYLSAVQPLNISQVFE-----VDTQSGVLSDRIR- 1021  
DB 864 -----EMYYLHLKKEENHIKLEEQKQITGTSTSSPGNTTNTVAQSAATHNSQNOQ 917  
QY 1022 -----TLATRIHELPLSL-----QDLTGLEHMLI--NCSKM--LPAD 1054  
DB 918 SNASSTNTQNGAVSSGPAVVEESH-PLTVLSISNDLKGIVSLNLGNKTKVPNPLTIS 976  
QY 1055 ITQL-----NNIPPTQESYIDPNLPVTKSLVTNCKPVT-----DKIHKAYKDK- 1098  
DB 977 TTEKKEFYENLLKNDYFNDDIKQFYKS---NSKVITGLTETOKNALNDEI-KKLDTL 1032  
QY 1099 -----NKYRFEI-----MGBEETAFKMTIRTNVSHVVGQLDIRKNRKFVCLND 1142  
DB 1033 QLSFDLYNKYKLDRLFNKKELQDQMKQIKKLTLLKEQLESKLSNL--NPNHN-VLQNF 1090  
QY 1143 NIDHNH 1149  
DB 1091 SVFFNKK 1097  
RESULT 3  
NTC2\_MOUSE STANDARD; PRT; 2470 AA.  
ID NTC2\_MOUSE  
AC O3516; Q60941; Q60608;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch B).  
DE NOTCH2.  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Thymus;  
RA Hamada Y., Higuchi M., Tsujimoto Y.;  
RT "Complete amino acid sequence and multiform transcripts encoded by a single copy of mouse Notch2 gene.";  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 316-1518 FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
RA MEDLINE=93178563; PubMed=8440332;  
RX Lardelli M., Lendahl U.;  
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues.";  
RL Exp. Cell Res. 204:364-372(1993).  
RN [3]  
RP SEQUENCE OF 1765-2153 FROM N.A.  
RX MEDLINE=97075110; PubMed=8917536;  
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D., Martin D.I.

RT "Inhibition of granulocytic differentiation by mNotch1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).  
RN [4].  
RP FUNCTION.  
RX MEDLINE=99396706; PubMed=10393120;  
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,  
RA Tsujimoto Y.;  
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early  
RT embryonic lethality.";  
RL Development 126:3415-3424(1999).  
RN [5].  
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.  
RX MEDLINE=95333893; PubMed=7609614;  
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;  
RT "Differential expression of Notch1 and Notch2 in developing and adult  
RT mouse brain.";  
RL Brain Res. Mol. Brain Res. 29:263-272(1995).  
RN [6].  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RX MEDLINE=21523956; PubMed=11518718;  
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (NL-4) undergo presenilin-dependent  
RT proteolysis.";  
RL J. Biol. Chem. 276:40268-40273(2001).  
RN [7].  
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.  
RX MEDLINE=21374376; PubMed=11459941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).  
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC Rap-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation.  
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytical processing NICD is translocated to the nucleus.  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
CC produced by alternative splicing.  
CC -!- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,  
CC neuroepithelia, somites, optic vesicles and branchial arches, but  
CC not heart.  
CC -!- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,  
CC the postnatal ependymal cells, and the choroid plexus throughout  
CC embryonic and postnatal development.  
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (TACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC -!- PTM: Phosphorylated.  
CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
CC -!- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.  
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration -  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D32210; BAA22094.1; -;  
CC EMBL; X68279; CAA48340.1; -;  
CC EMBL; U31881; AAC52924.1; -;  
CC HSSP; P16109; 1FSB.  
CC MGD; MGI:97364; Notch2.  
CC InterPro; IPR002110; ANK.  
CC InterPro; IPR000152; Asx\_hydroxyl.  
CC InterPro; IPR000561; EGF-like.  
CC InterPro; IPR000742; EGF\_2.  
CC InterPro; IPR001881; EGF\_Ca.  
CC InterPro; IPR001438; EGF\_II.  
CC InterPro; IPR000800; Notch.  
CC Pfam; PF00008; EGF; 35.  
CC Pfam; PF00023; ank; 6.  
CC Pfam; PF00066; notch; 2.  
CC PRINTS; PR00010; EGFBL00D.  
CC PRINTS; PR01452; NOTCH.  
CC SMART; SM00248; ANK; 4.  
CC SMART; SM00179; EGF\_CA; 22.  
CC SMART; SM00001; EGF\_Like; 12.  
CC SMART; SM00004; NL; 3.  
CC PROSITE; PS00088; ANK\_REPEAT; 4.  
CC PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
CC PROSITE; PS00010; ASX\_HYDROXYL; 22.  
CC PROSITE; PS00022; EGF\_1; 33.  
CC PROSITE; PS01186; EGF\_2; 27.  
CC PROSITE; PS01187; EGF\_CA; 22.  
CC Receptor; Transcription regulation; Activator; Differentiation;  
CC Developmental protein; Repeat; ANK repeat; EGF-like domain;  
CC Transmembrane; Glycoprotein; Signal; Phosphorylation;  
CC Alternative splicing.  
CC SIGNAL 1 25  
CC CHAIN 26 2470 POTENTIAL.  
CC CHAIN 1666 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
CC CHAIN 1697 2470 NOTCH EXTRACELLULAR TRUNCATION.  
CC DOMAIN 26 1677 NOTCH INTRACELLULAR DOMAIN.  
CC TRANSMEM 1678 1698 EXTRACELLULAR (POTENTIAL).  
CC DOMAIN 1699 2470 POTENTIAL.  
CC DOMAIN 26 63 CYTOPLASMIC (POTENTIAL).  
CC DOMAIN 65 102 EGF-LIKE 1.  
CC DOMAIN 105 143 EGF-LIKE 2.  
CC DOMAIN 144 180 EGF-LIKE 3.  
CC DOMAIN 182 219 EGF-LIKE 4.  
CC DOMAIN 221 256 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 296 334 EGF-LIKE 6 (INCOMPLETE).  
CC DOMAIN 336 372 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 373 411 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 413 452 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 454 490 EGF-LIKE 10.  
CC DOMAIN 492 528 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 530 566 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 568 603 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 605 641 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 643 678 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 680 716 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 718 753 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 755 791 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 793 829 EGF-LIKE 19.  
CC DOMAIN 831 869 EGF-LIKE 20.  
CC DOMAIN 871 907 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 909 945 EGF-LIKE 22.  
CC DOMAIN 947 983 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 985 1021 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1023 1059 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1061 1097 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1099 1145 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).  
CC DOMAIN 1147 1183 EGF-LIKE 28.  
CC EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).  
CC EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).







RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RL elegans.;  
 RL Nature 368:32-38(1994).  
 CC -!- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT  
 CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12  
 CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES  
 CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.  
 CC -!- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M12069; AAA70191.1; -;  
 DR EMBL; Z14092; CAA78474.1; -;  
 DR PIR; S06434; S06434;  
 DR HSSP; P00740; IEDM.  
 DR WormPep; R107.8; CE00274.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR000800; Notch.  
 DR Pfam; PF00008; EGF; 13.  
 DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00066; notch; 3.  
 DR PRINTS; PRO1452; NOTCH.  
 DR SMART; SM00248; ANK; 4.  
 DR SMART; SM00179; EGF\_CA; 2.  
 DR SMART; SM00001; EGF\_like; 10.  
 DR SMART; SM00004; NL; 2.  
 DR PROSITE; PS50088; ANK\_REPEAT; 3.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE; PS00022; EGF\_1; 12.  
 DR PROSITE; PS01186; EGF\_2; 11.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 1429  
 FT DOMAIN 16 908  
 FT TRANSMEM 909 931  
 FT DOMAIN 932 1429  
 FT DOMAIN 20 61  
 FT DOMAIN 114 150  
 FT DOMAIN 152 190  
 FT DOMAIN 201 246  
 FT DOMAIN 250 285  
 FT DOMAIN 287 323  
 FT DOMAIN 323 363  
 FT DOMAIN 365 402  
 FT DOMAIN 404 441  
 FT DOMAIN 449 492  
 FT DOMAIN 503 541  
 FT DOMAIN 543 579  
 FT DOMAIN 582 619  
 FT REPEAT 635 669  
 FT REPEAT 670 710  
 FT REPEAT 711 750

REPEAT 1093 1122 ANK 1.  
 REPEAT 1126 1158 ANK 2.  
 REPEAT 1162 1194 ANK 3.  
 REPEAT 1206 1236 ANK 4.  
 REPEAT 1240 1269 ANK 5.  
 FT DISULFID 24 35 BY SIMILARITY.  
 FT DISULFID 29 49 BY SIMILARITY.  
 FT DISULFID 51 60 BY SIMILARITY.  
 FT DISULFID 118 129 BY SIMILARITY.  
 FT DISULFID 123 138 BY SIMILARITY.  
 FT DISULFID 140 149 BY SIMILARITY.  
 FT DISULFID 156 169 BY SIMILARITY.  
 FT DISULFID 163 178 BY SIMILARITY.  
 FT DISULFID 180 189 BY SIMILARITY.  
 FT DISULFID 205 227 BY SIMILARITY.  
 FT DISULFID 221 234 BY SIMILARITY.  
 FT DISULFID 236 245 BY SIMILARITY.  
 FT DISULFID 254 264 BY SIMILARITY.  
 FT DISULFID 275 273 BY SIMILARITY.  
 FT DISULFID 275 284 BY SIMILARITY.  
 FT DISULFID 291 302 BY SIMILARITY.  
 FT DISULFID 296 311 BY SIMILARITY.  
 FT DISULFID 313 322 BY SIMILARITY.  
 FT DISULFID 327 339 BY SIMILARITY.  
 FT DISULFID 334 351 BY SIMILARITY.  
 FT DISULFID 353 362 BY SIMILARITY.  
 FT DISULFID 369 381 BY SIMILARITY.  
 FT DISULFID 375 390 BY SIMILARITY.  
 FT DISULFID 392 401 BY SIMILARITY.  
 FT DISULFID 408 419 BY SIMILARITY.  
 FT DISULFID 413 429 BY SIMILARITY.  
 FT DISULFID 431 440 BY SIMILARITY.  
 FT DISULFID 507 518 BY SIMILARITY.  
 FT DISULFID 512 529 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 547 558 BY SIMILARITY.  
 FT DISULFID 552 567 BY SIMILARITY.  
 FT DISULFID 569 578 BY SIMILARITY.  
 FT DISULFID 586 597 BY SIMILARITY.  
 FT DISULFID 591 607 BY SIMILARITY.  
 FT DISULFID 609 618 BY SIMILARITY.  
 FT CARBOHYD 41 41 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1429 AA; 157115 MW; 2355EDD7A62C025DB CRC64;  
 Query Match 2.6%; Score 166.5; DB 1; Length 1429;  
 Best Local Similarity 19.8%; Pred. No. 0.08;  
 Matches 111; Conservative 43; Mismatches 166; Indels 241; Gaps 21;  
 Qy 428 CAEGCPGSGWIKDGYCDKACNNACDWGDCSGNSGSRVYAGGGTSGVGPWQ--- 484  
 Db 643 CSERA-----NDGNCADACNYACKFDGDCSGKR-----EPFSKCR 679  
 Qy 485 -----FGGINSVSYCNQGCAN-----SWLADKFC 510  
 Db 680 YGNMCADFFANGV-----CNOACNNECLVDGMDCLPAVVRCVPVKTRHCASFANGIC 734  
 Qy 511 QACNVLCGFDAGDCGDHFEHLYKVILLNQTHYIIPKGECLPYFSFAEVAKRGEVAY 570  
 Db 735 PECNTNCGFDGDC-----DNET----- 753  
 Qy 571 SDNPPIIRHASTANKWKTIHLIMHSGMNAIT--HFNLTFTONTDEEFK----- 616  
 Db 754 -----NATIITNIRITVQ-MDPKEFQVGGQSILMTESSA 786  
 Qy 617 MQITVEYD-TREGPKLNSTAQKGYENLVSPITL-----LPEAEILFEDIPEKRFKPKRH 671

```

Db 787 LRVTVRIORDEGLV---FQWNGESEMVRKMKNERQLTEQHVLTSTISR-----KIKRS 838
Qy 672 DVNSTRQAQEVKIPLVNISLLPKDAQLSNTLDLQLEHGDITLKGYNLS-----KS 723
Db 839 ATNIGVVVILEVQENCDTGKCLKDAQSVVDSISARLAKKGDIDFGIPISALVAEPRKS 898
Qy 724 ALLRSFL-----MNSQHAQIKN 740
Db 899 GNNTGFLSNALLIGAGCLIVVVLMLGALPGNRTKRKRMINASVYMPMEENEKRNKN 958
Qy 741 QAITDETNDLSIAPAQKV-----HKSILPNSLGVSE-----RLQRLTTP 781
Db 959 HQSITSSQHSLLSEASYDGYIKRQNELQHYSLYPNPQGYGNGNDFLGDFNHTNLQIPTTP 1018
Qy 782 AVSVKVNHGDPQGNPLDLETTARFVE-----THTOKT---TGGNVTKEKPPSLIIVPLES 834
Db 1019 EPSPKILHTEAGSVAITEPTITRESVNIIDPRHNTVLTWLTASNSAEKSEDLIV---- 1074
Qy 835 QMTKEKKITGKEKENSMEEN 855
Db 1075 HEAKETIAAGADVNMDCDEN 1095

RESULT 6
NTC2_HUMAN STANDARD; PRT; 2471 AA.
AC Q04721; Q9H240; Q99734;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (hn2).
GN NCTCH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
RA Blaumueeller C.M., Mann R.S.;
RT "Complete human notch 2 (hn2) cDNA sequence.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
TISSUE=Breast tumor;
RA Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
RT "Human Notch2, a novel member of cell-fate determining NOTCH family.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 967-1229 FROM N.A.
TISSUE=T-cell;
RA Lemasson I., Devaux C., Mesnard J.M.;
RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE OF 1810-2447 FROM N.A.
TISSUE=Brain;
RX MEDLINE=93265135; PubMed=1303260;
RA Stifani S., Blaumueeller C.M., Redhead N.J., Hill R.E.,
RA Artavanis-Tsakonas S.;
RT "Human homologs of a Drosophila enhancer of split gene product define a novel family of nuclear proteins.";
RL Nat. Genet. 2:119-127(1992).
[5]
POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=97386453; PubMed=9244302;
RA Blaumueeller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane.";
RL Cell 90:281-291(1997).
[6]

```

```

RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794(1999).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
CC skeletal muscle and liver.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TMP-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -!- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF308601; AAA36377.2; -
DR EMBL; AF315356; AAG37073.1; -
DR EMBL; U77493; AAB19224.1; -
DR HSSP; P00740; IEDM.
DR Genew; HGNC:7882; NOTCH2.
DR MIN; 600275; -
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00248; ANK; 4.
DR SMART; SM00179; EGF_CA; 22.
DR SMART; SM00001; EGF_Like; 12.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.

```

DR PROSITE; PS01187; EGF\_CA; 22.  
 KW Receptor; Transcription regulation; Activator; Differentiation;  
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 2471 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.  
 FT CHAIN 1666 2471 NOTCH EXTRACELLULAR TRUNCATION (BY  
 FT CHAIN 1697 2471 SIMILARITY).  
 FT CHAIN 1697 2471 NOTCH INTRACELLULAR DOMAIN (BY  
 FT CHAIN 1697 2471 SIMILARITY).  
 FT DOMAIN 26 1677 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1678 1698 POTENTIAL.  
 FT DOMAIN 1699 2471 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 26 63 EGF-LIKE 1.  
 FT DOMAIN 64 102 EGF-LIKE 2.  
 FT DOMAIN 105 143 EGF-LIKE 3.  
 FT DOMAIN 144 180 EGF-LIKE 4.  
 FT DOMAIN 182 219 EGF-LIKE 5.  
 FT DOMAIN 221 258 EGF-LIKE 6.  
 FT DOMAIN 260 296 EGF-LIKE 7.  
 FT DOMAIN 298 336 EGF-LIKE 8.  
 FT DOMAIN 338 374 EGF-LIKE 9.  
 FT DOMAIN 375 413 EGF-LIKE 10.  
 FT DOMAIN 415 454 EGF-LIKE 11.  
 FT DOMAIN 456 492 EGF-LIKE 12.  
 FT DOMAIN 494 530 EGF-LIKE 13.  
 FT DOMAIN 532 568 EGF-LIKE 14.  
 FT DOMAIN 570 605 EGF-LIKE 15.  
 FT DOMAIN 607 643 EGF-LIKE 16.  
 FT DOMAIN 645 680 EGF-LIKE 17.  
 FT DOMAIN 682 718 EGF-LIKE 18.  
 FT DOMAIN 720 755 EGF-LIKE 19.  
 FT DOMAIN 757 793 EGF-LIKE 20.  
 FT DOMAIN 795 831 EGF-LIKE 21.  
 FT DOMAIN 833 871 EGF-LIKE 22.  
 FT DOMAIN 873 909 EGF-LIKE 23.  
 FT DOMAIN 911 947 EGF-LIKE 24.  
 FT DOMAIN 949 985 EGF-LIKE 25.  
 FT DOMAIN 987 1023 EGF-LIKE 26.  
 FT DOMAIN 1025 1061 EGF-LIKE 27.  
 FT DOMAIN 1063 1099 EGF-LIKE 28.  
 FT DOMAIN 1101 1147 EGF-LIKE 29.  
 FT DOMAIN 1149 1185 EGF-LIKE 30.  
 FT DOMAIN 1187 1223 EGF-LIKE 31.  
 FT DOMAIN 1225 1262 EGF-LIKE 32.  
 FT DOMAIN 1264 1302 EGF-LIKE 33.  
 FT DOMAIN 1304 1343 EGF-LIKE 34.  
 FT DOMAIN 1374 1412 EGF-LIKE 35.  
 FT REPEAT 1420 1456 LIN/NOTCH 1.  
 FT REPEAT 1503 1535 LIN/NOTCH 2.  
 FT REPEAT 1827 1871 ANK 1.  
 FT REPEAT 1876 1905 ANK 2.  
 FT REPEAT 1909 1939 ANK 3.  
 FT REPEAT 1943 1972 ANK 4.  
 FT REPEAT 1976 2005 ANK 5.  
 FT REPEAT 2009 2038 ANK 6.  
 FT DOMAIN 1645 1648 POLY-ALA.  
 FT DOMAIN 1994 1997 POLY-SER.  
 FT DOMAIN 2426 2429 BY SIMILARITY.  
 FT DISULFID 28 41 BY SIMILARITY.  
 FT DISULFID 35 51 BY SIMILARITY.  
 FT DISULFID 53 62 BY SIMILARITY.  
 FT DISULFID 68 79 BY SIMILARITY.  
 FT DISULFID 73 90 BY SIMILARITY.  
 FT DISULFID 92 101 BY SIMILARITY.  
 FT DISULFID 109 121 BY SIMILARITY.  
 FT DISULFID 115 131 BY SIMILARITY.  
 FT DISULFID 133 142 BY SIMILARITY.  
 FT DISULFID 148 159 BY SIMILARITY.  
 FT DISULFID 153 168 BY SIMILARITY.  
 FT DISULFID 170 179 BY SIMILARITY.  
 FT DISULFID 186 198 BY SIMILARITY.  
 FT DISULFID 192 207 BY SIMILARITY.

FT DISULFID 209 218 BY SIMILARITY.  
 FT DISULFID 225 236 BY SIMILARITY.  
 FT DISULFID 230 246 BY SIMILARITY.  
 FT DISULFID 248 257 BY SIMILARITY.  
 FT DISULFID 264 275 BY SIMILARITY.  
 FT DISULFID 269 284 BY SIMILARITY.  
 FT DISULFID 286 295 BY SIMILARITY.  
 Query Match 2.68; Score 166.5; DB 1; Length 2471;  
 Best Local Similarity 19.7%; Pred. No. 0.18;  
 Matches 110; Conservative 64; Mismatches 176; Indels 207; Gaps 26;  
 QY 424 PVPNCAECPCGSKWIKDGYCDKACNNSACDWDGDCSGNSGSRVAGGGTSGVGPW 483  
 DB 1421 PPATCLSQYCADKARDGVGCDEACNSHACQWDGDC-----SUTMENPW 1463  
 QY 484 -QFGGGINSVSYCQGC-----ANS-----WLADKF-----CDQACN 514  
 DB 1464 ANCSPLPCWDYINNQCDELNTVECLDFNFCOGNSKTKYKYCADHFKDNHCNOGN 1523  
 QY 515 VLSGCFDAGCGQDHFEHLYK-----VILLPNOTHYIIPKGECLPYFSFAEV----- 561  
 DB 1524 SEECGWDGLDCAADQPENLAEGTLVIVVLM-----PEQLQDARSFLRALTLLHTN 1576  
 QY 562 -AKRGVGGAYSDNPIIRHASIANKWKTIHLIMHSGMNATTIHFNLTFONTNDEPK--- 616  
 DB 1577 LRIKRSOGELMVYIYGEKSAAMKQ-----RMTRRSLPGQEQEVA 1619  
 QY 617 -MQITVEVDTRGPKLNSTAQYENLVSPITLLPEAEILFEDIPKEKRPFKRRHDVNS 675  
 DB 1620 GSKVFEIDNRQ---CVQSDSHCFKNTDAAALLAS-----HAIQG 1657  
 QY 676 TRRAQEEVKIPLVNI---SLLPKDAQSLNTLDLQ-----EHGDI 714  
 DB 1658 T-----LSYPLSVSVSESLTPERTQL-LYLLAVAVIILFILLGVIMAKRKHGSLW 1710  
 QY 715 L-KGVNLSKALLRSFLMNSOHAK-----IKNOAIITDENDSLVAPQEKVHKS 763  
 DB 1711 LPEGFTLRDA-----SNHKRRPVGQDAVGLKNLSVOVSEAN--LI----- 1750  
 QY 764 ILPNSLGVSERLQRLTFPAVSVKYNHGDQ---QNPPDLDETARFRVETHTKTIGN 819  
 DB 1751 ---GTGTFSEHWVDEGPQPK-KVKADEALLSEEDPIDR-----RPWTQOHEAA 1797  
 QY 820 VTKKPPSLVPLESOMTKE-----KKTGKEKENSMEENAEH---H 859  
 DB 1798 DIRRTPSLALTPQAEQEVLDVNVGRPGDCTPLMLASLRGGSDLSDEDAEDSSAN 1857  
 QY 860 IGVTEVLLGRKLQHYTD 876  
 DB 1858 IITDLIVQGASLQATD 1874  
 RESULT 7  
 NTCL\_RAT  
 ID NTCL\_RAT STANDARD; PRT; 2531 AA.  
 AC Q07008;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).  
 GN NOTCH1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N A.  
 RC TISSUE-Schwann cell;  
 RX MEDLINE=92111383; PubMed=1764995;  
 RA Weinmaster G., Roberts V.J., Lemke G.;  
 RT "A homolog of Drosophila Notch expressed during mammalian  
 development.";

Development 113:199-205(1991).  
 [2] CC  
 RP REVISIONS TO 1652-1653.  
 RA Weinmaster G.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 [3] CC  
 RN FUNCTION.  
 RP MEDLINE=21094508; PubMed=11182080;  
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,  
 RA Honjo T.;  
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent  
 RT neural progenitor cells to an astroglial fate."  
 RL Neuron 29:45-55(2001).  
 [4] CC  
 RN TISSUE SPECIFICITY.  
 RP MEDLINE=93202015; PubMed=1295745;  
 RA Weinmaster G., Roberts V.J., Lenke G.;  
 RT "Notch2: a second mammalian Notch gene."  
 RL Development 116:931-941(1992).  
 [5] CC  
 RN TISSUE SPECIFICITY.  
 RP MEDLINE=21331789; PubMed=11438922;  
 RA Irvin D.K., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;  
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple  
 RT functional roles for the Notch-DSL signaling system during brain  
 RT development."  
 RL J. Comp. Neurol. 436:167-181(2001).  
 CC  
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands  
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
 CC Upon ligand activation through the released notch intracellular  
 CC domain (NICD) it forms a transcriptional activator complex with  
 CC RBP-J kappa and activates genes of the enhancer of split locus.  
 CC Affects the implementation of differentiation, proliferation and  
 CC apoptotic programs (By similarity). Acts instructively to control  
 CC the cell fate determination of CNS multipotent progenitor cells,  
 CC resulting in astroglial induction and neuron/oligodendrocyte  
 CC suppression.  
 CC  
 CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
 CC terminal fragment N(EC) which are probably linked by disulfide  
 CC bonds (By similarity).  
 CC  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following  
 CC proteolytical processing NICD is translocated to the nucleus (By  
 CC similarity).  
 CC  
 CC -!- TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.  
 CC Expressed in postnatal central nervous system (CNS) germinal zones  
 CC and, in early postnatal life, within numerous cells throughout the  
 CC CNS. Found in both subventricular and ventricular germinal zones.  
 CC  
 CC -!- DEVELOPMENTAL STAGE: In the embryo, highest levels occur between  
 CC days 12 and 14 and decrease rapidly to much lower levels in the  
 CC adult.  
 CC  
 CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
 CC which is proteolytically cleaved by a furin-like convertase in the  
 CC trans-Golgi network before it reaches the plasma membrane to yield  
 CC an active, ligand-accessible form. Cleavage results in a C-  
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
 CC (FACE) to yield a membrane-associated intermediate fragment called  
 CC notch extracellular truncation (NEXT). This fragment is then  
 CC cleaved by presenilin dependent gamma-secretase to release a  
 CC notch-derived peptide containing the intracellular domain (NICD)  
 CC from the membrane (By similarity).  
 CC  
 CC -!- PTM: Phosphorylated (By similarity).  
 CC  
 CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
 CC  
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC  
 CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 -----  
 CC EMBL; X57405; CAA040667.1; -  
 CC HSSP; P00740; 1EDM.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF 2.  
 DR InterPro: IPR001881; EGF.Ca.  
 DR InterPro: IPR001438; EGF II.  
 DR InterPro: IPR002049; Laminin\_EGF.  
 DR InterPro: IPR000800; Notch.  
 DR Pfam: PF00008; EGF; 36.  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF00066; notch; 3.  
 DR PRINTS; PR00010; EGFBL00D.  
 DR PRINTS; PR00011; EGFAMININ.  
 DR PRINTS; PR01452; NOTCH.  
 DR SMART; SM00248; ANK; 5.  
 DR SMART; SM00179; EGF\_CA; 25.  
 DR SMART; SM00001; EGF\_Like; 10.  
 DR SMART; SM00004; NL; 2.  
 DR PROSITE; PS50088; ANK\_REPEAT; 4.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 35.  
 DR PROSITE; PS01186; EGF\_2; 26.  
 DR PROSITE; PS01187; EGF\_CA; 21.  
 KW Receptor; Transcription regulation; Activator; Differentiation;  
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.  
 FT SIGNAL 1 18  
 FT CHAIN 19 2531 NOTCH EXTRACELLULAR TRUNCATION (BY  
 FT CHAIN 1711 2531 SIMILARITY).  
 FT CHAIN 1744 2531 NOTCH INTRACELLULAR DOMAIN (BY  
 FT SIMILARITY).  
 FT DOMAIN 19 1723 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1724 1746 POTENTIAL.  
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 20 58 EGF-LIKE 1.  
 FT DOMAIN 59 99 EGF-LIKE 2.  
 FT DOMAIN 102 139 EGF-LIKE 3.  
 FT DOMAIN 140 176 EGF-LIKE 4.  
 FT DOMAIN 178 216 EGF-LIKE 5.  
 FT DOMAIN 218 255 EGF-LIKE 6.  
 FT DOMAIN 257 293 EGF-LIKE 7.  
 FT DOMAIN 295 333 EGF-LIKE 8.  
 FT DOMAIN 335 371 EGF-LIKE 9.  
 FT DOMAIN 372 410 EGF-LIKE 10.  
 FT DOMAIN 412 450 EGF-LIKE 11.  
 FT DOMAIN 452 488 EGF-LIKE 12.  
 FT DOMAIN 490 526 EGF-LIKE 13.  
 FT DOMAIN 528 564 EGF-LIKE 14.  
 FT DOMAIN 566 603 EGF-LIKE 15.  
 FT DOMAIN 603 639 EGF-LIKE 16.  
 FT DOMAIN 641 676 EGF-LIKE 17.  
 FT DOMAIN 678 714 EGF-LIKE 18.  
 FT DOMAIN 716 751 EGF-LIKE 19.  
 FT DOMAIN 753 789 EGF-LIKE 20.  
 FT DOMAIN 791 827 EGF-LIKE 21.  
 FT DOMAIN 829 867 EGF-LIKE 22.  
 FT DOMAIN 869 905 EGF-LIKE 23.  
 FT DOMAIN 907 943 EGF-LIKE 24.  
 FT DOMAIN 945 981 EGF-LIKE 25.  
 FT DOMAIN 983 1019 EGF-LIKE 26.  
 FT DOMAIN 1021 1057 EGF-LIKE 27.  
 FT DOMAIN 1059 1095 EGF-LIKE 28.  
 FT DOMAIN 1097 1143 EGF-LIKE 29.  
 FT DOMAIN 1145 1181 EGF-LIKE 30.  
 FT DOMAIN 1183 1219 EGF-LIKE 31.  
 FT DOMAIN 1221 1265 EGF-LIKE 32.  
 FT DOMAIN 1267 1305 EGF-LIKE 33.

FT DOMAIN 1307 1346 EGF-LIKE 34.  
 FT DOMAIN 1348 1384 EGF-LIKE 35.  
 FT DOMAIN 1387 1426 EGF-LIKE 36.  
 FT REPEAT 1445 1480 LIN/NOTCH 1.  
 FT REPEAT 1481 1522 LIN/NOTCH 2.  
 FT REPEAT 1523 1562 LIN/NOTCH 3.  
 FT REPEAT 1917 1946 ANK 1.  
 FT REPEAT 1950 1980 ANK 2.  
 FT REPEAT 1984 2013 ANK 3.  
 FT REPEAT 2017 2046 ANK 4.  
 FT REPEAT 2050 2079 ANK 5.  
 FT DOMAIN 1730 1733 POLY-ALA.  
 FT DOMAIN 1891 1894 POLY-GLU.  
 FT DOMAIN 2258 2261 POLY-PRO.  
 FT DOMAIN 2497 2500 POLY-SER.  
 FT SITE 1654 1655 CLEAVAGE BY URIN-LIKE PROTEASE (BY  
 FT DISULFID 24 37 BY SIMILARITY.  
 FT DISULFID 31 46 BY SIMILARITY.  
 FT DISULFID 48 57 BY SIMILARITY.  
 FT DISULFID 63 74 BY SIMILARITY.  
 FT DISULFID 68 87 BY SIMILARITY.  
 FT DISULFID 89 98 BY SIMILARITY.  
 FT DISULFID 106 117 BY SIMILARITY.  
 FT DISULFID 111 127 BY SIMILARITY.  
 FT DISULFID 129 138 BY SIMILARITY.  
 FT DISULFID 144 155 BY SIMILARITY.  
 FT DISULFID 149 164 BY SIMILARITY.  
 FT DISULFID 166 175 BY SIMILARITY.  
 FT DISULFID 182 195 BY SIMILARITY.  
 FT DISULFID 189 204 BY SIMILARITY.  
 FT DISULFID 206 215 BY SIMILARITY.  
 FT DISULFID 222 233 BY SIMILARITY.  
 FT DISULFID 227 243 BY SIMILARITY.  
 FT DISULFID 245 254 BY SIMILARITY.  
 FT DISULFID 261 272 BY SIMILARITY.  
 Query Match 2.6%; Score 164; DB 1; Length 2531;  
 Best Local Similarity 19.8%; Pred. No. 0.26;  
 Matches 109; Conservative 61; Mismatches 171; Indels 210; Gaps 27;  
 QY 402 GKDWPDYFYSKGVKYLTPVPCAECPGSGWIKGYCDKACNNSACDWDGDCSGN 461  
 DB 1437 GRDIPP-----PQIEACELPCEQDAGNK-----VCNLCNNHACGWDGDC--- 1479  
 QY 462 SGRSRYIAGGGTSGVGPQWQ-----FGGINSYSYCNQ-GC----- 499  
 DB 1480 -----SLNFDPMKNCQSLQCKWYFSDG-HCDSQCNACGCLFDGDCQLT 1524  
 QY 500 -----ANSWLADKFCQACNLSCGFDAGCGODHPHE-----LYKVILLP-- 540  
 DB 1525 EGOCNPLYDQYKDHFDHSGCHGCGNACSEWDGLDCA-EHYPERLAAGTIVLVLLPPD 1583  
 QY 541 -----NOTHYIIPKGECL-----PYFSFAEVAKRGVEGAYSNDPIIRHA 579  
 DB 1584 QLRNNSFHLRDYSHVLTHTNVVFKRDAQOQQMIFPY-----GREELRKHPIKRS 1635  
 QY 580 SIANKWKTTHLI-----MHSGMATTTHFNLTQNTNDEEFKQIIVVDVRE----- 627  
 DB 1636 V---GWATTSLLPGTNGGRQRELDPMDDHGSIVY-----LEIDNROCVOSS 1679  
 QY 628 -----GPKLNSTAGQYENLVSPITLLPEAAILPEDIP----- 660  
 DB 1680 SQCFQSATDVAFLGALASLSINIPYKLEAV-KSETVEPPLPSQLHLMYVAAAFAVLLF 1738  
 QY 661 -----KKR-----FKRHDVNSTRAQAEVKIPL--VNISLLP-KDAQLS 700  
 DB 1739 FVCGVLLSRKRRRQHQWLFPEGFKVSEASKRRE-----PLGDSVGLKPLKNASDG 1793  
 QY 701 LNTLDLQLEHG--DITLKYNLSKALLRSFLMNSQHKAKIKQAIIITDETNDLSVAPQEK 758  
 DB 1794 ALMDDNQNEWGEDLETKKFRFEPEPVLPLDLDQDTHRQWTOHQHDAADLRVSAMAPTTP 1853

QY 759 QVHKSILPNSLGVSRQLRTPFAVSKVYKNGHOGQNPDLDETARPRVETHQKTIGG 818  
 DB 1854 Q-----GEVDADCMNVNVRGPD--GFTPLMIASCSGGLET-----GN 1889  
 QY 819 NVTKKEKPPSLI 829  
 DB 1890 SEEEEDAPAVI 1900  
 RESULT 8  
 NTCL\_BRARE STANDARD; PRT; 2437 AA.  
 AC P46530;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Neurogenic locus notch homolog protein 1 precursor.  
 GN NOTCH1A OR NOTCH.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OC NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA MEDLINE=94128602; PubMed=8297791;  
 RA Bierkamp C., Campos-Ortega J.A.;  
 RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and  
 RL its pattern of transcription during early embryogenesis."; Mech. Dev. 43:87-100(1993).  
 CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING OF THE  
 CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE  
 CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation  
 CC stages. During gastrulation is differentially expressed, accumulating predominantly in the prechordal mesoderm and notochord. At the end of gastrulation, expressed along the anterior-posterior axis including the developing neural plate and differentiating mesoderm. Also present in the developing brain and head regions.  
 CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X69088; CAA48831.1; -  
 CC HSPSP; P00740; IEDM.  
 CC ZFIN; ZDB-GENE-990415-173; notchla.  
 CC InterPro; IPR002110; ANK.  
 CC InterPro; IPR000152; Asx\_hydroxyl.  
 CC InterPro; IPR000561; EGF-like.  
 CC InterPro; IPR000742; EGF\_2.  
 CC InterPro; IPR001881; EGF\_Ca.  
 CC InterPro; IPR001438; EGF\_II.  
 CC InterPro; IPR000800; Notch.  
 CC Pfam; PF00008; EGF; 36.  
 CC Pfam; PF00023; ank; 6.  
 CC Pfam; PF00066; notch; 3.  
 CC PRINTS; PR00010; EGFBL00D.  
 CC PRINTS; PR01452; NOTCH.  
 CC SMART; SM00248; ANK; 5.  
 CC SMART; SM00179; EGF\_Ca; 19.  
 CC SMART; SM00001; EGF\_like; 16.

DR SMART; SM00004; NL; 3.  
DR PROSITE; PS00088; ANK\_REPEAT; 4.  
DR PROSITE; PS00297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00100; ASX\_HYDROXYL; 23.  
DR PROSITE; PS00022; EGF\_1; 34.  
DR PROSITE; PS01186; EGF\_2; 28.  
DR PROSITE; PS01187; EGF\_CA; 22.  
KW Receptor; Transcription regulation; Activator; Differentiation;  
KW Developmental protein; Neurogenesis; Repeat; ANK repeat;  
KW EGF-like domain; Transmembrane; Glycoprotein; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.  
FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1725 1747 POTENTIAL.  
FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 21 57 EGF-LIKE 1.  
FT DOMAIN 58 98 EGF-LIKE 2.  
FT DOMAIN 101 138 EGF-LIKE 3.  
FT DOMAIN 139 175 EGF-LIKE 4.  
FT DOMAIN 177 215 EGF-LIKE 5.  
FT DOMAIN 217 254 EGF-LIKE 6.  
FT DOMAIN 256 292 EGF-LIKE 7.  
FT DOMAIN 294 332 EGF-LIKE 8.  
FT DOMAIN 334 370 EGF-LIKE 9.  
FT DOMAIN 371 409 EGF-LIKE 10.  
FT DOMAIN 411 449 EGF-LIKE 11.  
FT DOMAIN 451 487 EGF-LIKE 12.  
FT DOMAIN 489 524 EGF-LIKE 13.  
FT DOMAIN 526 562 EGF-LIKE 14.  
FT DOMAIN 564 599 EGF-LIKE 15.  
FT DOMAIN 601 637 EGF-LIKE 16.  
FT DOMAIN 639 674 EGF-LIKE 17.  
FT DOMAIN 676 712 EGF-LIKE 18.  
FT DOMAIN 714 749 EGF-LIKE 19.  
FT DOMAIN 751 787 EGF-LIKE 20.  
FT DOMAIN 789 825 EGF-LIKE 21.  
FT DOMAIN 827 865 EGF-LIKE 22.  
FT DOMAIN 867 903 EGF-LIKE 23.  
FT DOMAIN 905 941 EGF-LIKE 24.  
FT DOMAIN 943 979 EGF-LIKE 25.  
FT DOMAIN 981 1017 EGF-LIKE 26.  
FT DOMAIN 1019 1055 EGF-LIKE 27.  
FT DOMAIN 1057 1093 EGF-LIKE 28.  
FT DOMAIN 1095 1141 EGF-LIKE 29.  
FT DOMAIN 1143 1179 EGF-LIKE 30.  
FT DOMAIN 1181 1217 EGF-LIKE 31.  
FT DOMAIN 1219 1263 EGF-LIKE 32.  
FT DOMAIN 1265 1303 EGF-LIKE 33.  
FT DOMAIN 1305 1344 EGF-LIKE 34.  
FT DOMAIN 1346 1382 EGF-LIKE 35.  
FT DOMAIN 1385 1423 EGF-LIKE 36.  
FT REPEAT 1446 1486 LIN/NOTCH 1.  
FT REPEAT 1487 1520 LIN/NOTCH 2.  
FT REPEAT 1521 1561 LIN/NOTCH 3.  
FT REPEAT 1867 1910 ANK 1.  
FT REPEAT 1915 1944 ANK 2.  
FT REPEAT 1948 1978 ANK 3.  
FT REPEAT 1982 2011 ANK 4.  
FT REPEAT 2015 2044 ANK 5.  
FT REPEAT 2048 2077 ANK 6.  
FT DOMAIN 2265 2276 POLY-GLN (OPA-REPEAT).  
FT SITE 1656 CLEAVAGE BY FURIN-LIKE PROTEASE (BY  
SIMILARITY).  
FT DISULFID 25 35 BY SIMILARITY.  
FT DISULFID 29 45 BY SIMILARITY.  
FT DISULFID 47 56 BY SIMILARITY.  
FT DISULFID 62 73 BY SIMILARITY.  
FT DISULFID 67 86 BY SIMILARITY.  
FT DISULFID 88 97 BY SIMILARITY.  
FT DISULFID 105 116 BY SIMILARITY.  
FT DISULFID 118 126 BY SIMILARITY.  
FT DISULFID 128 137 BY SIMILARITY.  
FT DISULFID 143 154 BY SIMILARITY.

FT DISULFID 148 163 BY SIMILARITY.  
FT DISULFID 165 174 BY SIMILARITY.  
FT DISULFID 181 194 BY SIMILARITY.  
FT DISULFID 188 203 BY SIMILARITY.  
FT DISULFID 205 214 BY SIMILARITY.  
FT DISULFID 221 232 BY SIMILARITY.  
FT DISULFID 226 242 BY SIMILARITY.  
FT DISULFID 244 253 BY SIMILARITY.  
FT DISULFID 260 271 BY SIMILARITY.  
FT DISULFID 265 280 BY SIMILARITY.  
FT DISULFID 282 291 BY SIMILARITY.  
FT DISULFID 298 311 BY SIMILARITY.  
FT DISULFID 305 320 BY SIMILARITY.  
FT DISULFID 322 331 BY SIMILARITY.  
FT DISULFID 338 349 BY SIMILARITY.  
FT DISULFID 343 358 BY SIMILARITY.  
FT DISULFID 360 369 BY SIMILARITY.  
FT DISULFID 375 386 BY SIMILARITY.  
FT DISULFID 380 397 BY SIMILARITY.  
FT DISULFID 399 408 BY SIMILARITY.  
FT DISULFID 415 428 BY SIMILARITY.  
FT DISULFID 422 437 BY SIMILARITY.  
FT DISULFID 439 448 BY SIMILARITY.  
FT DISULFID 455 466 BY SIMILARITY.  
FT DISULFID 460 475 BY SIMILARITY.  
FT DISULFID 477 486 BY SIMILARITY.  
FT DISULFID 493 503 BY SIMILARITY.  
FT DISULFID 498 512 BY SIMILARITY.  
FT DISULFID 514 523 BY SIMILARITY.  
FT DISULFID 530 541 BY SIMILARITY.  
FT DISULFID 535 550 BY SIMILARITY.  
FT DISULFID 552 561 BY SIMILARITY.  
FT DISULFID 568 578 BY SIMILARITY.  
FT DISULFID 573 587 BY SIMILARITY.  
FT DISULFID 589 598 BY SIMILARITY.  
FT DISULFID 605 616 BY SIMILARITY.  
FT DISULFID 625 636 BY SIMILARITY.  
FT DISULFID 627 636 BY SIMILARITY.  
FT DISULFID 643 653 BY SIMILARITY.  
FT DISULFID 648 662 BY SIMILARITY.  
FT DISULFID 664 673 BY SIMILARITY.  
FT DISULFID 680 691 BY SIMILARITY.  
FT DISULFID 685 700 BY SIMILARITY.  
FT DISULFID 702 711 BY SIMILARITY.  
FT DISULFID 718 728 BY SIMILARITY.  
FT DISULFID 723 737 BY SIMILARITY.  
FT DISULFID 739 748 BY SIMILARITY.  
FT DISULFID 755 766 BY SIMILARITY.  
FT DISULFID 760 775 BY SIMILARITY.  
FT DISULFID 777 786 BY SIMILARITY.  
FT DISULFID 793 804 BY SIMILARITY.  
FT DISULFID 798 813 BY SIMILARITY.  
FT DISULFID 815 824 BY SIMILARITY.  
FT DISULFID 831 842 BY SIMILARITY.  
FT DISULFID 836 853 BY SIMILARITY.  
FT DISULFID 855 864 BY SIMILARITY.  
FT DISULFID 871 882 BY SIMILARITY.  
FT DISULFID 876 891 BY SIMILARITY.  
FT DISULFID 893 902 BY SIMILARITY.  
FT DISULFID 909 920 BY SIMILARITY.  
FT DISULFID 914 929 BY SIMILARITY.  
FT DISULFID 931 940 BY SIMILARITY.  
FT DISULFID 947 958 BY SIMILARITY.  
FT DISULFID 952 967 BY SIMILARITY.  
FT DISULFID 969 978 BY SIMILARITY.

Query Match 2.68; Score 163.5; DB 1; Length 2437;

Best Local Similarity 20.7%; Pred. No. 0.26; Indels 197; Gaps 29;  
Matches 119; Conservative 58; Mismatches 201;Qy 411 YSHSKGQ-----KVLTPVPCARGCPGCKGCKACNNACDWDGDCGNS 462  
||| ||| : : : ||| : ||| ||| ||| |||  
Db 1427 YFSGGQGRDIAPPVEVEIRCIAC-EGRGNAI----CDTCNNHACGWDGDC---- 1477





Db 2130 K-----NEVSIKALQVYGNRIFTDKNGKKGISFFIPIKODTLKTIINETI 2181  
Qy 811 -----HTOKTIGNVTK-----EKPPSLIVPLESQMTKEKKITCKEENSME 853  
Db 2182 DNSNIKQRLGIYIFVRKNVSNSEKLCDFTTGTSMLNELSQV-KEKCKTVKIKKGDIFG 2240  
Qy 854 ENAENHIGV-----TEVLLGRKLQHYTDS-YLGFPLWEKKY-----FDDLDEEBSL 900  
Db 2241 LKCPKGAIFPQACSFNSVLLVEYKSDYEDSEHINYIHKDKKYNLKPDKVIELMDEN--- 2297  
Qy 901 KQLAYFTDSKNRARKR-DTFADSLRVN-KILNSKFGFTSRKVPAAH--MPHMIDRIVM 956  
Db 2298 -----FRELQIQOYTGISNITDVLHPKFNGLNPLNFKNHYSTAYAKVPDTFNSIIN 2351  
Qy 957 QELQDMRPE-----EFDKTSFKVHRHSEDMQFAP-----SYFYLLMSAVOPLNIS 1001  
Db 2352 FSCNCPNPEKHVYQTMQVESNRPNDIKNKNWIKNELLNIEKYALLDDEERQKKIK 2411  
Qy 1002 QVFDEVDTDQSGVLSDRIRTLATRIHELPLSLQDLTGLEHMLI----- 1045  
Db 2412 QQEEEQEQE-ILKQDDR-----LSRHDDYKNKHTYILYDSNEHICDYKNES 2459  
Qy 1046 -----NCSKMLPADITOLN-----NIPPTQESYDNPPLPVPVTSKLVNMC 1084  
Db 2460 LISTLPNDTKKIQKSICKINAKALDVTIKCPHTKNETPKD---YFPN-----SSLITND 2511  
Qy 1085 KPVTDKIHK-----AYKDKNRYFEI 1105  
Db 2512 KKVITFDKKNFVYIIDTKTFSL 2536

RESULT 10  
RA50\_METJA STANDARD; PRT; 1005 AA.  
ID RA50\_METJA  
AC Q58718;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA double-strand break repair rad50 AtPase.  
GN RAD50 OR MJ1322.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337993; PubMed=868087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glode A.,  
RA Scott J.L., Geohagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii".  
RL Science 273:1058-1073(1996).  
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
CC rad50/mrell complex possesses single-strand endonuclease activity  
CC and ATP-dependent double-strand-specific exonuclease activity.  
CC Rad50 provides an ATP-dependent control of mrell by unwinding  
CC and/or repositioning DNA ends into the mrell active site (By  
CC similarity).  
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; U67572; AAB99331.1; -.  
DR TIGR; MJ1322; -.  
DR InterPro; IPR003439; ABC\_transportr.  
DR InterPro; IPR003405; SMC\_C.  
DR InterPro; IPR003395; SMC\_N.  
DR Pfam; PF02463; SMC\_N; 1.  
DR Pfam; PF02483; SMC\_C; 1.  
DR ProDom; PD000006; ABC\_transportr; 1.  
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
FT NP\_BIND 32 39 ATP (BY SIMILARITY).  
FT DOMAIN 158 849 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E78F3 CRC64;  
  
Query Match 2.6%; Score 162.5; DB 1; Length 1005;  
Best Local Similarity 19.2%; Pred. No. 0.081;  
Matches 139; Conservative 124; Mismatches 255; Indels 205; Gaps 33;  
  
Qy 593 NKWKTIIHLSGNNATTIHEN-LTFQNTDEEFKMQITVEVDTRREGPKLNSTAQKGYEN 641  
Db 232 NEWENKKLLYEKFINKEERKRALEKQELKILEYDINTVVEARETUNRRHDEYEKYS 291  
Qy 642 LVSEPT-----TLLPEAEILFEDIPKEKRF-PKFK-RHVDVNSTRAQE 681  
Db 292 LVDEIRKTESRLRELKSHYEDYLKLTQLEIKGIDIEKLEFINKSKYRDDIDNLDLTLN 351  
Qy 682 EVKIPLVNISLLPKDAQLSLNTLDLQLEHGDITLKGYNLSKALLRSFLMSQHAQIKNQ 741  
Db 352 KIKDEIERVETI-KDLLEELKNLEEIE---KIEKYKRICECKEY---EKYLEBEK 403  
Qy 742 AIIITDE-TNDSLVAPOEKQVHKSLIPNSLGVSERLORLTPFAVSVKVNGHOGQNPPLDL 800  
Db 404 AVEYNKLTLEVTILLOEK---KSTKKNINDLETRINKLL-----BETKN--IDI 447  
Qy 801 ETTARFRVETHQKTIGNVTKRPP--SLIVPLESQMTKEKKITGKEK----- 847  
Db 448 ESIENSLKEIEKKKVLLENQKEKTELKNGEINSEIKRLKILDELKEVEGKCPCKT 507  
Qy 848 ---ENSRMEENAEHIGVTEVLLGRKLQHYTDSYLGFLPWEKKYFIDL-----LDE 896  
Db 508 PIDENKME-----LINQHTQLANKYTELEBEINKKIREIEKDEKLEKKEIDK 555  
Qy 897 EESLKTQLAYETDSKNR---ARYKRDTPADSLRVYVK-----ILNSKFGFTSRKVPAAHP 948  
Db 556 EENLTKLTLYLEKQSQIELELELKNYKEQLDEINKKISNVYNGK----- 602  
Qy 949 HMIDRVNQELQDMFPEEDFTSFHKVHRHSEDMQFASFYFYLLMSAVOPLN-----ISQ 1002  
Db 603 -PVDEI-----LEDI-----KSQNLNFK-----NFYQYLSAVSYLSNVDEEGRN 642  
Qy 1003 VFDEVDTDQSG-----VLSDEIRTLATRIHELPLSLQDLTGLEH-MLINS 1048  
Db 643 RIKEIENIVGWNNKEKREELNKLREDEIRNLKDLKNELEKKEKLEIENIRSLKFD 702  
Qy 1049 KW-----LPADITQINNPPTQESYDNPPLPVTKSLVNTCKPVT-DKIHKAYDKNRYR 1102  
Db 703 KYKEVLGLTEKLEELKNKIDGLEIYN-----ICNSKILAIIDNKKRY---NKED 749  
Qy 1103 FEIWEDEEI-----AFKMRITNVSHVVGQLDDIRKNP---RKVCLNDN----- 1143  
Db 750 IEIVLNNKILEVNKEINDIEERISVINGKLEIDINYNEEKKIKELYNKQELDNVREQ 809  
Qy 1144 -----IDHNHDKAOTVKAVLROFYESMFFIPSOQELPREYRNRFLLHHELQEWRAVD 1196  
Db 810 KTEIETGLEYLKKOVESLAKRKEM-----SNLEKEKEKLTKEVE-----YID 852  
Qy 1197 KIK 1199  
Db 853 KVR 855

RESULT 11  
NTCL\_MOUSE  
ID NTCL\_MOUSE STANDARD; PRT; 2531 AA.  
AC 001705; Q9QW58; Q99JC2; Q06007; Q61905; Q9R0X7;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Notch A)  
GN NOTCH1 OR NOTCH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Embryo;  
RX MEDLINE=93194170; PubMed=8449489;  
RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,  
RA Copeland N.G., Gridley T.;  
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse  
RT homolog of Drosophila Notch.";  
RL Genomics 15:259-264(1993).  
RN [2]  
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.  
RC STRAIN=CD-1; TISSUE=Embryo;  
RX MEDLINE=93050801; PubMed=1426644;  
RA Reaume A.G., Conlon R.A., Ziringibi R., Yamaguchi T.P., Rossant J.;  
RT "Expression analysis of a Notch homologue in the mouse embryo.";  
RL Dev. Biol. 154:377-387(1992).  
RN [3]  
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.  
RC TISSUE=Embryo;  
RX MEDLINE=93048835; PubMed=1425352;  
RA Greenspan R.J., McMahon A.P., Gridley T.;  
RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,  
RT suggests an important role in early postimplantation mouse  
RT development.";  
RL Development 115:737-744(1992).  
RN [4]  
RP SEQUENCE OF 1161-1547 FROM N.A.  
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;  
RX MEDLINE=93178563; PubMed=8440332;  
RA Lardelli M., Lendahl U.;  
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a  
RT wide variety of tissues.";  
RL Exp. Cell Res. 204:364-372(1993).  
RN [5]  
RP SEQUENCE OF 1659-1673 FROM N.A.  
RX MEDLINE=99364499; PubMed=10437788;  
RA Lee J.S., Ishimoto A., Yanagawa S.I.;  
RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads  
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";  
RL FEBS Lett. 455:276-280(1999).  
RN [6]  
RP SEQUENCE OF 1950-2201 FROM N.A.  
RX MEDLINE=98029496; PubMed=9384671;  
RA Messierle M., Follo M., Nehls M., Eggert H., Boehm T.;  
RT "Dynamic changes in gene expression during in vitro differentiation of  
RT mouse embryonic stem cells.";  
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).  
RN [7]  
RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND  
RP MUTAGENESIS OF 1651-ARG-GLN-ARG-ARG-1654.  
RX MEDLINE=98318619; PubMed=9653148;  
RA Logeat F., Bessia C., Brou C., Lebail O., Jarriault S., Seidah N.G.,  
RA Israel A.;  
RT "The Notch1 receptor is cleaved constitutively by a furin-like  
RT convertase.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).  
RN [8]

PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.  
RX MEDLINE=21523956; PubMed=11518718;  
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;  
RT "Murine notch homologs (NL-4) undergo presenilin-dependent  
RT proteolysis.";  
RN J. Biol. Chem. 276:40268-40273(2001).  
RN [9]  
RP POST-TRANSLATIONAL PROCESSING.  
RX MEDLINE=21374376; PubMed=11459941;  
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;  
RT "Conservation of the biochemical mechanisms of signal transduction  
RT among mammalian Notch family members.";  
RN Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).  
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands  
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.  
CC Upon ligand activation through the released notch intracellular  
CC domain (NICD) it forms a transcriptional activator complex with  
CC RBP-J kappa and activates genes of the enhancer of split locus.  
CC Affects the implementation of differentiation, proliferation and  
CC apoptotic programs (By similarity). May play an essential role in  
CC postimplantation development, probably in some aspect of cell  
CC specification and/or differentiation. May be involved in mesoderm  
CC development, somite formation and neurogenesis. Involved in the  
CC maturation of both CD4+ and CD8+ cells in the thymus.  
CC -I- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-  
CC terminal fragment N(EC) which are probably linked by disulfide  
CC bonds.  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Following  
CC proteolytic processing NICD is translocated to the nucleus.  
CC -I- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
CC produced by alternative splicing.  
CC -I- TISSUE SPECIFICITY: Highly expressed in the brain, lung and  
CC thymus. Expressed at lower levels in the spleen, bone-marrow,  
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal  
CC muscle, kidney and heart.  
CC -I- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 d.p.c.  
CC By 8.5 d.p.c. highly expressed in presomitic mesoderm, mesenchyme  
CC and endothelial cells, while much lower levels are seen in the  
CC neuroepithelium. Between 9.5-10.5 d.p.c. expressed at high levels  
CC in the neuroepithelium. At 13.5 d.p.c. expressed in the surface  
CC ectoderm, eye and developing whisker follicles.  
CC -I- PTM: Synthesized in the endoplasmic reticulum as an inactive form  
CC which is proteolytically cleaved by a furin-like convertase in the  
CC trans-Golgi network before it reaches the plasma membrane to yield  
CC an active, ligand-accessible form. Cleavage results in a C-  
CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following  
CC ligand binding, it is cleaved by TNF-alpha converting enzyme  
CC (TACE) to yield a membrane-associated intermediate fragment called  
CC notch extracellular truncation (NEXT). This fragment is then  
CC cleaved by presenilin dependent gamma-secretase to release a  
CC notch-derived peptide containing the intracellular domain (NICD)  
CC from the membrane.  
CC -I- PHOSPHORYLATION.  
CC -I- SIMILARITY: BELONGS TO THE NOTCH FAMILY.  
CC -I- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
CC -I- SIMILARITY: CONTAINS 5 LIN/NOTCH REPEATS.  
CC -I- SIMILARITY: CONTAINS 5 ANK REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z11886; CAA77941.1; -;  
DR EMBL; L02613; AAK14898.1; -;  
DR EMBL; X68278; CAA48339.1; -;  
DR EMBL; AJ238029; CAB40733.1; -;  
DR EMBL; X82562; CAA57909.1; -;  
DR HSSP; P00740; IEDM.  
DR MGI; 97363; Notchl1.

DR	InterPro: IPR002110; ANK.	
DR	InterPro: IPR000152; Asx_hydroxyl.	
DR	InterPro: IPR000561; EGF-like.	
DR	InterPro: IPR000742; EGF-2.	
DR	InterPro: IPR001881; EGF-Ca.	
DR	InterPro: IPR001438; EGF-11.	
DR	InterPro: IPR000800; Notch.	
DR	Pfam: PF00008; EGF; 35.	
DR	Pfam: PF00023; ank; 7.	
DR	Pfam: PF00066; notch; 3.	
DR	PRINTS: PR00010; EGFBL00D.	
DR	PRINTS: PR01452; NOTCH.	
DR	SMART: SM00248; ANK; 3.	
DR	SMART: SM00179; EGF-Ca; 23.	
DR	SMART: SM00001; EGF-like; 11.	
DR	SMART: SM00004; NL; 2.	
DR	PROSITE: PS50088; ANK_REPEAT; 2.	
DR	PROSITE: PS50297; ANK_REPEAT_REGION; 1.	
DR	PROSITE: PS00010; ASX_HYDROXYL; 22.	
DR	PROSITE: PS00022; EGF-1; 34.	
DR	PROSITE: PS01186; EGF-2; 27.	
DR	PROSITE: PS01187; EGF-Ca; 21.	
KW	Receptor; Transcription regulation; Activator; Differentiation;	
KW	Developmental protein; Repeat; ANK repeat; EGF-like domain;	
KW	Transmembrane; Glycoprotein; Signal; Phosphorylation;	
KW	Alternative splicing.	
FT	SIGNAL 1 18	POTENTIAL.
FT	CHAIN 19 2531	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT	CHAIN 1711 2531	NOTCH EXTRACELLULAR TRUNCATION.
FT	CHAIN 1744 2531	NOTCH INTRACELLULAR DOMAIN.
FT	DOMAIN 1719 1725	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1726 1746	POTENTIAL.
FT	DOMAIN 1747 2531	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 20 58	EGF-LIKE 1.
FT	DOMAIN 59 99	EGF-LIKE 2.
FT	DOMAIN 102 139	EGF-LIKE 3.
FT	DOMAIN 140 176	EGF-LIKE 4.
FT	DOMAIN 178 216	EGF-LIKE 5.
FT	DOMAIN 218 255	EGF-LIKE 6.
FT	DOMAIN 257 293	EGF-LIKE 7.
FT	DOMAIN 295 333	EGF-LIKE 8.
FT	DOMAIN 335 371	EGF-LIKE 9.
FT	DOMAIN 372 410	EGF-LIKE 10.
FT	DOMAIN 412 450	EGF-LIKE 11.
FT	DOMAIN 452 488	EGF-LIKE 12.
FT	DOMAIN 490 526	EGF-LIKE 13.
FT	DOMAIN 528 564	EGF-LIKE 14.
FT	DOMAIN 566 601	EGF-LIKE 15.
FT	DOMAIN 603 639	EGF-LIKE 16.
FT	DOMAIN 641 676	EGF-LIKE 17.
FT	DOMAIN 678 714	EGF-LIKE 18.
FT	DOMAIN 716 751	EGF-LIKE 19.
FT	DOMAIN 753 789	EGF-LIKE 20.
FT	DOMAIN 791 827	EGF-LIKE 21.
FT	DOMAIN 829 867	EGF-LIKE 22.
FT	DOMAIN 869 905	EGF-LIKE 23.
FT	DOMAIN 907 943	EGF-LIKE 24.
Query Match	2.6%;	Score 162; DB 1; Length 2531;
Best Local Similarity	19.7%;	Pred. No. 0.34;
Matches 117; Conservative	29.78; 66;	Mismatches 196; Indels 216; Gaps 29;
QY	424 PVPNCAEGC--PGSWIKDG--YCDKACNNSACDWDGDCSGNSGGSRYIAGGGTGSIGV 479	-----SLNF 1483
Db	1441 PPTPEEACELPECOVDAGNKVCNLCNNHACGWDGDC-----	-----
QY	480 GQPWQ-----FGGGINSVSYCQ--GC-----	-----ANSWLAD 506
Db	1484 NDPWNKCTQSLQCKWKYFSDG--HCDSCNCSAGCLFDGFCQLTGCQNPLYDQYCKDHFS 1542	-----
QY	507 KFCDOACNVLSGGFDAGDCGDHFHE-----LYKVLLLP-----NQTHYIIPGECLP 554	-----
Db	1543 GHCDGCGNSACEWDLGLDCA--EHVPERLAGTILVLYLPDQLRNNSHFFL---RELISH 1598	-----

```
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS00096; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Alkylation; Multigene family.
FT DOMAIN 1 778
FT DOMAIN 779 808
FT DOMAIN 809 841
FT DOMAIN 841 1927
FT NP_BIND 174 181
FT DOMAIN 654 676
FT MOD_RES 694
FT MOD_RES 704 704
FT SEQUENCE 1961 AA; 226336 MW; 9B9876D9681F19E CRC64;
Query Match 2.5%; Score 158.5; DB 1; Length 1961;
Best Local Similarity 21.6%; Pred. No. 0.38;
Matches 121; Conservative 104; Mismatches 246; Indels 89; Gaps 23;
QY 622 EVDTRGPKLNSTAGQYENLVSPITLLPEAEILFEDIPKEKRPKFKRHVDVNSTRRQAE 681
DQ 622 EVDTRGPKLNSTAGQYENLVSPITLLPEAEILFEDIPKEKRPKFKRHVDVNSTRRQAE 681
DB 870 EMTWOSQLMAEKLQ-DEQOAKTELCAEAEELRATKQOELEETCHDLEA-RVSEE 927
QY 682 EYKIPLVNISLPPKDAQSLNLTDLQLEHGDITLKGYNLSK---SALLR-----SFLMNS 733
DQ 682 EYKIPLVNISLPPKDAQSLNLTDLQLEHGDITLKGYNLSK---SALLR-----SFLMNS 733
DB 928 EERCQYLQAE---KKMQQNIQLEQLBEESARQKQLEKVTYPAKLLKLEEQIIMED 985
QY 734 QHAKI-KNOAIITD---ETNDSLVAPQEKVQHSILPNS-----LGVSERLQRTFFPAVS 784
DQ 734 QHAKI-KNOAIITD---ETNDSLVAPQEKVQHSILPNS-----LGVSERLQRTFFPAVS 784
DB 986 QNCKLAKEKLEEDRAVEFTTDLMEEEKSKSLAKLNKHEAMITDLERLRR-----1038
QY 785 VKVNGHDGQNPPLDLETTARFVETHQKTIGGNVTKPKPSLIVPLESQMTEKKITG 844
DQ 785 VKVNGHDGQNPPLDLETTARFVETHQKTIGGNVTKPKPSLIVPLESQMTEKKITG 844
DB 1039 -----EEKQROELE-----KTRKLEGDSTDLN---DQAELOAQIAELKMQLA 1079
QY 845 KEKEN-----SRMEEN-AENHIGVTEVLGRKLOHYTDSYLGLPWEK-----KKYFLD 892
DQ 845 KEKEN-----SRMEEN-AENHIGVTEVLGRKLOHYTDSYLGLPWEK-----KKYFLD 892
DB 1080 KKEEELQALARAVEEAQKNALKKI---RELETQISELOEDLESEACRNKAEKQRD 1136
QY 893 LDEEESLQTLAYTDSKNRARYKRDFTADSLRVYVKNILNSKFGFTSRKVPVAPHMID 952
DQ 893 LDEEESLQTLAYTDSKNRARYKRDFTADSLRVYVKNILNSKFGFTSRKVPVAPHMID 952
DB 1137 LGEELEALKTELEDTLDTAAQOELRSKREQEVSLKKTLEDE---AKTHEAQIQEM--1190
QY 953 RIVMQELODMFPPEFDKTSFHKVRHSDMOPAFSYFYILMSAVOPL-----NISQV 1003
DQ 953 RIVMQELODMFPPEFDKTSFHKVRHSDMOPAFSYFYILMSAVOPL-----NISQV 1003
DB 1191 ROKHSQAVELEAEQLEQTKRVKATLEKAKQTLNENGERELANEVKALLOGKGDSEHKRVV 1250
QY 1004 FDEVDTDQSGVLSDEIRIT-LATRIHELPLSLQDLTGLEHMLINGSKMLPADITOL-NNI 1061
DQ 1004 FDEVDTDQSGVLSDEIRIT-LATRIHELPLSLQDLTGLEHMLINGSKMLPADITOL-NNI 1061
DB 1251 EAQLQELQVFKFSEGERVTRADKYSKQLQVELDSVTGLLNQSDSKSKLTKDFSALQSOL 1310
QY 1062 PPTQESYDNPVPPVTKSLVNTCKPVTDKIHKAYDKKNKYRFEIMGEIEEIAFKMTRTNVS 1121
DQ 1062 PPTQESYDNPVPPVTKSLVNTCKPVTDKIHKAYDKKNKYRFEIMGEIEEIAFKMTRTNVS 1121
DB 1311 QDTQELQELLENKQKL--SLSTKLQMED-----EKNSFRQLEEEEEEAARNLEKQIA 1361
QY 1122 HVYQGLDDIRKPNKRV-CL 1140
DQ 1122 HVYQGLDDIRKPNKRV-CL 1140
DB 1362 TLHAQVTDMMKKMEDGVGCL 1381
RESULT 13
ID SBCC_CLOAB STANDARD; PRT; 1163 AA.
AC Q97FK1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclease sbcCD subunit C.
GN SBCC OR CAC2736.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- FUNCTION: sbcCD cleaves DNA hairpin structures. These structures
CC can inhibit DNA replication and are intermediates in certain DNA
CC recombination reactions. The complex acts as a 3'->5' double
CC strand exonuclease that can open hairpins. It also has a 5'
CC single-strand endonuclease activity (by similarity).
CC -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE007771; AK80682.1; -.
CC InterPro: IPR003439; ABC_transportr.
CC Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
CC DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 35 42
FT DOMAIN 197 415
FT DOMAIN 446 1003
FT SEQUENCE 1163 AA; 135507 MW; CE5F0BD2215D7A92 CRC64;
Query Match 2.5%; Score 158; DB 1; Length 1163;
Best Local Similarity 21.5%; Pred. No. 0.19;
Matches 129; Conservative 95; Mismatches 195; Indels 182; Gaps 31;
QY 630 KLNSTAGK-----YENLVSPITLLPEAEILFEDIPKEKRPKFKRHVDVNSTRRQAEV 683
DQ 630 KLNSTAGK-----YENLVSPITLLPEAEILFEDIPKEKRPKFKRHVDVNSTRRQAEV 683
DB 501 KLNDNRQKWKYSYNSLSKASLRVENSEQVLR--TKKEMTKL-----ED 545
QY 684 KIPLVNTSLPLPKDAQSLNLTDLQLEHG-----DITLKGYNLSKALL 726
DQ 684 KIPLVNTSLPLPKDAQSLNLTDLQLEHG-----DITLKGYNLSKALL 726
DB 546 KISKVNKISLETENNAHVLRKLSGGEACPGCVGSHHKEGFKVDLKAETLKE-L 604
QY 727 RSFLMNSQAHAKNQAIITDETNDLSVAPQEKVQHSILPNSL-----YSERLQRTFF 781
DQ 727 RSFLMNSQAHAKNQAIITDETNDLSVAPQEKVQHSILPNSL-----YSERLQRTFF 781
DB 605 EGF---EKRRFENEIVMCEASIKVEKNIKLINESI--NNLGEPEKVESMEKKFN 659
QY 782 AVSVKVGND-----QGNPPDL-L-ETTARFVETHQKTIGGNVTKPKPSLIVPLESQM 836
DQ 782 AVSVKVGND-----QGNPPDL-L-ETTARFVETHQKTIGGNVTKPKPSLIVPLESQM 836
DB 660 YLKEKVNKFNLEKIQDNDNKDLSERSNKIEVEYQKEKTEKQCKR-----IVDLKSEL 714
QY 837 TK-----EKKITG-----KEKNSRMEENAEHNVTEVLGRKL 871
DQ 837 TK-----EKKITG-----KEKNSRMEENAEHNVTEVLGRKL 871
DB 715 EEAKEFNEVAYTIENKAEKIQDFEMKEILEKRVVE--AGEIKDLNLL--NI 770
QY 872 QHYTDSYGLFVPEWKYFLDLDDEESLKTQLAYFTDSKNRARYKRDFTADSLRVYVKNKI 931
DQ 872 QHYTDSYGLFVPEWKYFLDLDDEESLKTQLAYFTDSKNRARYKRDFTADSLRVYVKNKI 931
DB 771 RH-----TEKEQ---LMDKCSRLKEEL-----SKNAELKEK-----DKI 802
QY 932 LNSKFGFTSRKVPV-----HMPHMDIRV-----MQELODMFPPEEDKTSFH 973
DQ 932 LNSKFGFTSRKVPV-----HMPHMDIRV-----MQELODMFPPEEDKTSFH 973
DB 803 INEKIELKKNVGVLDNLYELKEKIEGTIKIEQYNLCDDKMKNEIED----KYRCSDE 858
```

```

QY 974 KVRHSEDMQFAFYFYLMSAVQPLN---ISQVEDEVDTDQSGVLSREIRTLATRIHEL 1030
Db 859 IIKYHSNLS---SLKDRKVDIDKLNLKLMEEKFENIEKAKENYLNKLNKSDVEKY 915
QY 1031 PLSLODLTGLEHMLNSKMLPADITQINLPPTQESYYPDNPVTKYSLVTKNCKPVTDK 1090
Db 916 KNELSKVGAVEVL---SK-----KLKRNKLTTEEKWIE-----IQN-----NR 950
QY 1091 IHKAYKDKNKYRFEIMGEETI-----AFKMIRT--NVSHVVGOLDIDRK--NPRKF 1137
Db 951 VEKASKAKALQERSIKLEEVKNTEIKELKGLKLTQKLEHLKSLDLEKLFKGGKF 1010
QY 1138 V 1138
Db 1011 V 1011

RESULT 14
MYHB_CHICK
ID MYHB_CHICK STANDARD; PRT; 1978 AA.
AC P10587;
DT 01-JUL-1989 (Rel. 11, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, gizzard smooth muscle.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=88118918; PubMed=2892941;
RA Yanagisawa M., Hamada Y., Katsuragawa Y., Inamura M., Mikawa T.,
RA Masaki T.;
RT "Complete primary structure of vertebrate smooth muscle myosin heavy
RT chain deduced from its complementary DNA sequence. Implications on
RT topography and function of myosin.";
RL J. Mol. Biol. 198:143-157(1987).
RN [2]
RP REVISIONS.
RA Masaki T.;
RL Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE OF 1-203.
RX MEDLINE=88032919; PubMed=3312184;
RA Maita T., Onishi H., Yajima E., Matsuda G.;
RT "Amino acid sequence of the amino-terminal 24 kDa fragment of the
RT heavy chain of chicken gizzard myosin.";
RL J. Biochem. 102:133-145(1987).
RN [4]
X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-818, AND REVISIONS 204-215.
RX MEDLINE=98412652; PubMed=9741621;
RA Dominguez R., Freyzon Y., Trybus K.M., Cohen C.;
RT "Crystal structure of a vertebrate smooth muscle myosin motor domain
RT and its complex with the essential light chain: visualization of the
RT pre-power stroke state.";
RL Cell 94:559-571(1998).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).

```

```

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06546; CAA29793.1; -.
DR PIR; S03166; S03166.
DR PDB; 1BR1; 09-SEP-98.
DR PDB; 1BR2; 09-SEP-98.
DR PDB; 1BR4; 09-SEP-98.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ; 1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Calmodulin-binding;
FT MULTIGENE FAMILY; 3D-structure.
FT INIT_MET 0
FT DOMAIN 1 790 MYOSIN HEAD-LIKE.
FT DOMAIN 791 820 IQ.
FT DOMAIN 849 1978 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 849 1978 COILED COIL (POTENTIAL).
FT NP_BIND 176 183 ATP.
FT DOMAIN 666 688 ACTIN-BINDING.
FT DOMAIN 767 781 ACTIN-BINDING.
FT MOD_RES 1 1 BLOCKED.
FT MOD_RES 127 127 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 706 706 ALKYLATION (SH-1).
FT MOD_RES 716 716 ALKYLATION (SH-2).
FT CONFLICT 127 127 MISSING (IN REF. 3).
FT CONFLICT 204 215 KDTISITQGPSFS -> RTPASLVKHLPP (IN REF. 1).
SQ SEQUENCE 1978 AA; 228663 MW; B7B6C923E5273D93 CRC64;

Query Match 2.5%; Score 157.5; DB 1; Length 1978;
Best Local Similarity 20.8%; Pred. No. 0.44;
Matches 126; Conservative 122; Mismatches 250; Indels 109; Gaps 28;

QY 612 DEEFKMQITVEVDTFREGPKLNSTAGK-----YENLVSPITLLPEAELFEDIPK 661
Db 863 DEE--LQRTKEROQKAEAELEKELQKLTQCEKNLQKLAETELYAAEEMRVRLAA 920
QY 662 EKRPEKFRHDVNSIRRAQAEVKIPLVNISLLPKDAQLSLNLDLQLEHGDITLKGYNLS 721
Db 921 KQOLEEILHEWEARIEEERS---QOLQAEKKKMOOQMLDLEEQEEEAARQKLOLE 977
QY 722 KSALLRSFLMNSHAKIKNQAITDETNDLSVAPQ---EKQVHKSLIPSLGVSERLQR 778
Db 978 K-----VTADGKIKKMEDDILIMEDQNNKLTKEKLLEREV-SDLTINLAEEAKAKNL 1030
QY 779 TFPADVSVKVNHDGQGNPDLDL-----ETTARFRVETHTQKTIGCNVTKEKPPSLIVPLES 834
Db 1031 T-----KLKNKHESMISELEVRLLKKEKSKQELKIKRKEGSSDLHEQ---IAELQA 1081
QY 835 QMTKEKKITGKEKEN-----SRME-ENAEHNHIGVTEVLLGRKLQHYDYSYGLFLPWKE-- 886
Db 1082 QIAELKAQAKKEELQALARLEDETSQKNNAKKI---RELESHISDLQEDLESEKAA 1138

```



```

FT DOMAIN 771 808 EGF-LIKE 20.
FT DOMAIN 810 847 EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 849 885 EGF-LIKE 22. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 887 922 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 924 960 EGF-LIKE 24.
FT DOMAIN 962 998 EGF-LIKE 25.
FT DOMAIN 1000 1034 EGF-LIKE 26.
FT DOMAIN 1047 1082 EGF-LIKE 27.

Query Match 2.5%; Score 155.5; DB 1; Length 2321;
Best local Similarity 24.7%; Pred. No. 0.72;
Matches 38; Conservative 14; Mismatches 37; Indels 65; Gaps

QY 423 WVPNC-----AEGCPGSKT---DGYCDKACNNSACDWDGGDCSGNSGSGRY 467
Db 1367 WTGPCEAPAAPEVSEPRCPAAQAKRQDRCDCRENSPGCGWDGGDC----- 1417
QY 468 IAGGGGTGSGVGPQWQFGGGN-----SVSYCNQGCAN----- 501
Db 1418 -----SLSVGDPRQCEALQCHWLFNNSEKCDPACSSPACLIYDNFCHAGGRERCNP 1469

QY 502 ---SWLADF---CDQACNVLSGFDAGDCGD 528
: : : : :
Db 1470 VYEKICADHFADGRCDQGCNTEECGWDGLDCASE 1503

Search completed: April 13, 2003, 03:21:33
Job time : 53.9235 secs

```



STIC-Biotech/ChemLib

98589

From: Slobodyansky, Elizabeth  
Sent: Tuesday, April 01, 2003 3:25 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/023,888

Please search for case 10/023,888:

ES

SEQ ID NOs: 1, 2 and 22 against commercial and interference databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652  
CM1 10D11  
703-306-3222

mail box 10D01

Point of Contact:  
Mona Smith  
Technical Information Specialist  
CM1 6A01  
Tel: 308-3278

RECEIVED  
APR - 1 2003  
(STIC)

Searcher: M. Smith  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 4/9/03  
Date Completed: 4/14/03  
Searcher Prep/Review: 8  
Clerical: \_\_\_\_\_  
Online time: 12

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: 2  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

7

1. The first part of the document is a list of the names of the persons who were present at the meeting. The names are listed in alphabetical order.